SADA Sampler for Nonnegative Matrix Factorization

Onur Poyraz
Yusuf Hakan Kalaycı

Abstract

Space alternating data augmentation (SADA) proposed by Doucet et al. is an MCMC generalization of the space alternating generalized expectation-maximization (SAGE) algorithm. SADA Sampler is particularly well suited for models having a composite structure, i.e., when the data may be written as a sum of latent components. The SADA sampler is shown to have favorable mixing properties and lesser storage requirement when compared to standard Gibbs sampling. We will apply SADA algorithm to Nonnegative Matrix Factorization Models and we will compare the results of the Gibbs Sampler and SADA Sampler.

Introduction

The aim of this project is to implement an algorithm for nonnegative matrix factorization model. NMF can be considered as a sum of latent variables. For this types of models we can apply Gibbs Sampler to reproduce desired result. However, as an alternative to Gibbs Sampler, we used SADA Sampler which has good mixing properties and requires less storeage.

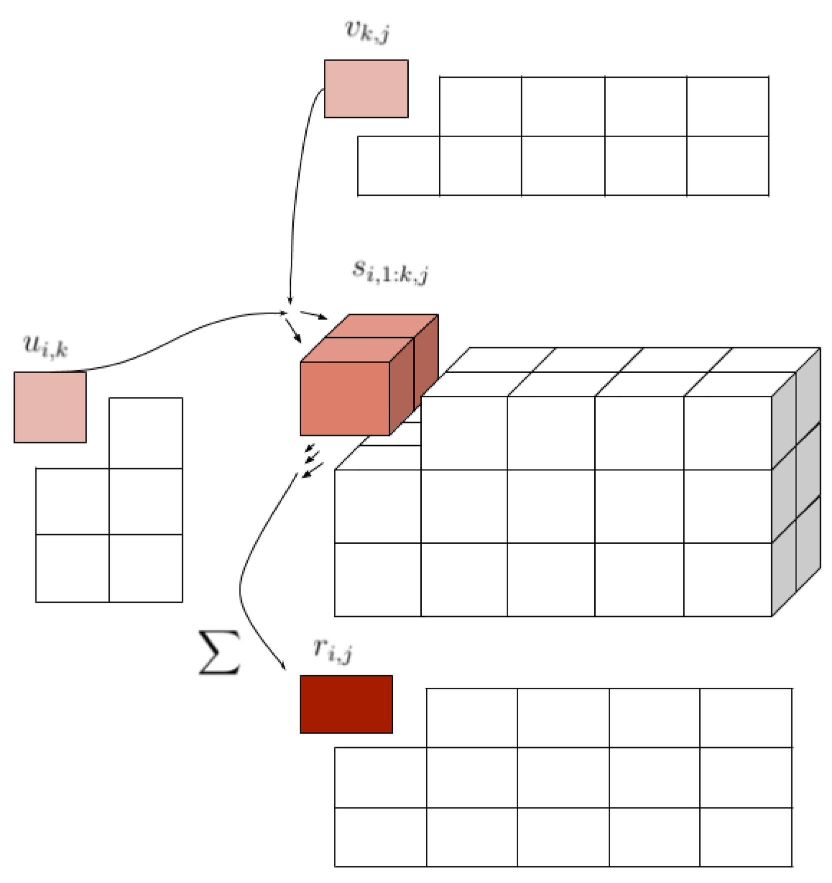
Problem

1 What is Nonnegative Matrix Factorization?

- Input $X = \begin{bmatrix} x_{11} & \dots & x_{1J} \\ \dots & \dots & \dots \\ x_{I1} & \dots & x_{IJ} \end{bmatrix}$
- Approach: $X \approx UV$
- Objective: $(U, V)^* = argmin_{U,V>0} D(X \| UV)$

Where $D(.\|.)$ represents a suitably chosen error function.

Model



$$u_{ik} \sim G(u_{ik}; A_{ik}^{(u)}, B_{ik}^{(u)})$$

$$v_{kj} \sim G(v_{kj}; A_{kj}^{(v)}, B_{kj}^{(v)})$$

$$s_{ikj} \sim PO(s_{ikj}; u_{ik} v_{kj})$$

$$X_{ij} \approx r_{ij} = \sum_k s_{ikj}$$

Derivation

Gibbs Sampler Full Conditionals

$$\mathbb{P}(s_{ikj}|u, v, X) = M(s_{ikj}; X_{ij}, \frac{u_i. * v_j}{u_i * v_j})$$

$$\mathbb{P}(u_{ik}|s, v, X, \theta) = G(u_{ik}; A_u + \sum_j s_{ikj}, B_u + M * v^T)$$

$$\mathbb{P}(v_{kj}|s, u, X, \theta) = G(v_{kj}; A_v + \sum_i s_{ikj}, B_v + u^T * M)$$

SADA Sampler Marginals

$$\mathbb{P}(s_{ij}^{(k)}|u^{(k)}, v^{(k)}, X) = BI(s_{ij}^{(k)}; X_{ij}, \frac{u_i k * v_k j}{u_i * v_j})$$

$$\mathbb{P}(u_i^{(k)}|s^{(k)}, v^{(k)}, \theta^{(k)}) = G(u_i^{(k)}; A_u^{(k)} + \sum_j s_{ikj}, B_u^{(k)} + M * {v^{(k)}}^T)$$

$$\mathbb{P}(v_j^{(k)}|s^{(k)}, u^{(k)}, \theta^{(k)}) = G(v_j^{(k)}; A_v^{(k)} + \sum_i s_{ikj}, B_v^{(k)} + {u^{(k)}}^T * M)$$

Proof of Convergence for SADA

Theorem 1. *Let $\pi(\theta_1, \dots, \theta_K)$ be target distribution. Assume that for each k there exists a latent variable c_k and a density q_k such that*

$$\int q_k(c_k, \theta_k, \theta_{-k}) dc_k = \pi(\theta_k, \theta_{-k})$$

then $c_k \sim q_k(c_k|\theta_k', \theta_{-k})$, $\theta_k \sim q_k(c_k, \theta_{-k})$ corresponds to a π-reversible move on coordinate θ_k.

Proof. The transition kernel from θ_k' to θ_k writes

$$\begin{aligned} K(\theta_k|\theta_k') &= \int q_k(\theta_k|c_k, \theta_{-k}) q_k(c_k|\theta_k', \theta_{-k}) dc_k \\ &= \int \frac{q_k(c_k, \theta_k, \theta_{-k})}{\int q_k(c_k, \widetilde{\theta}_k, \theta_{-k}) d\widetilde{\theta}_k} \frac{q_k(c_k, \theta_k', \theta_{-k})}{\pi(\theta_k', \theta_{-k})} dc_k \end{aligned} \tag{1}$$

and thus satisfies the detailed balance equation

$$K(\theta_k|\theta_k')\pi(\theta_k', \theta_{-k}) = K(\theta_k'|\theta_k)\pi(\theta_k, \theta_{-k}),$$

which indicates that $\pi(\theta_k, \theta_{-k})$ is stationary for $K(\theta_k|\theta_k')$. □

Results

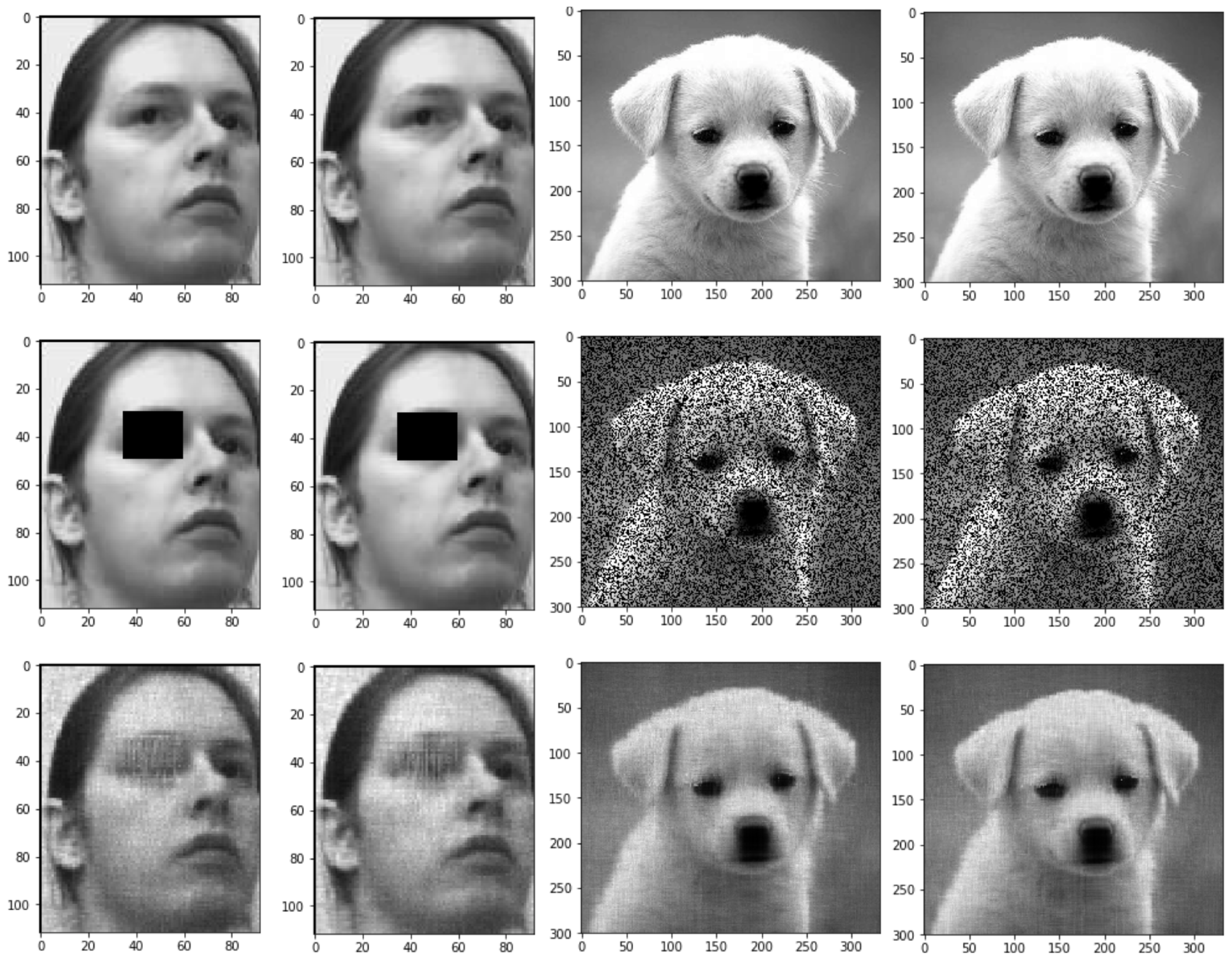


Fig.1 Comparison of the Gibbs Sampler and SADA Sampler outputs: The upper photos of each column are the original photo and the middle photos of each column correspond to the masked image which given to the samplers. The below photos of each column are the regenerated image from masked one. First and third column correspond to Gibbs while others correspond to SADA.
The SADA Sampler complete one epoch four times faster than Gibbs Sampler. However, Gibbs sampler requires less epoch for convergences. During the training, we see that since it requires more epochs still SADA Sampler is faster than Gibbs Sampler.

Conclusion

In this project, we implemented the composite model for NMF with using Gibbs and SADA sampler and compared the results. Although none of them is better in terms of divergence of the results, SADA sampler is faster and it requires lesser storage. Results of both Gibbs and SADA for image recovery are promising.

Reference

1. Fevotte et.al., Efficient MCMC Inference in Composite Models with SADA
2. Fevotte et al., Nonnegative Matrix Factorization as Probabilistic Inference in Composite Models
3. Umut Şimşekli et al., MCMC Inference for Probabilistic Latent Tensor Factorisation
4. Taylan Cemgil, Bayesian Inference for Nonnegative Matrix Factorization
5. $https://github.com/yhkalayci/SADA_Sampler$