UNIUNEA EUROPEANĂ

Instrumente Structurale 2014-2020

Metode de optimizare riemanniene pentru învăţare profundă
Proiect cofinanţat din Fondul European de Dezvoltare Regională prin Programul Operaţional Competitivitate 2014-2020

Variational Autoencoders Endowed with Richer but Still Computationally Efficient Statistical Models

Alexandra Peşte, **Luigi Malagò** {peste, malago}@rist.ro

Romanian Institute of Science and Technology

Abstract

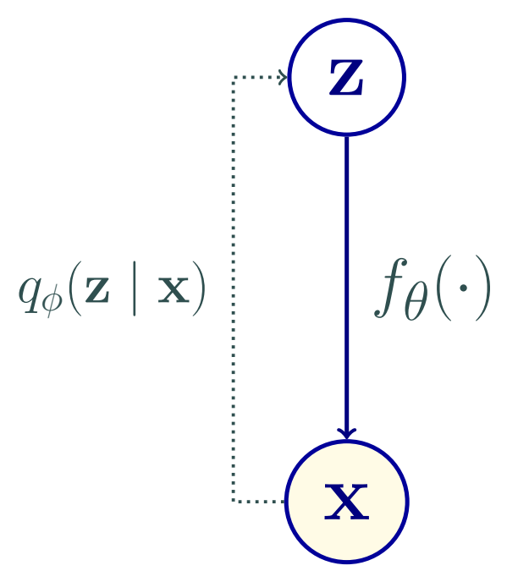
Variational autoencoders (VAEs) are one of the most powerful tools for approximate inference in the context of deep learning. We study the use of Gaussian Graphical Models for the approximation of both the posterior of the latent variables and the conditional distribution of the observations. Two examples are presented: a chain model and a regular grid, in order to capture correlations not represented by the independence model, allowing efficient stochastic backpropagation.

Background on Variational Autoencoders

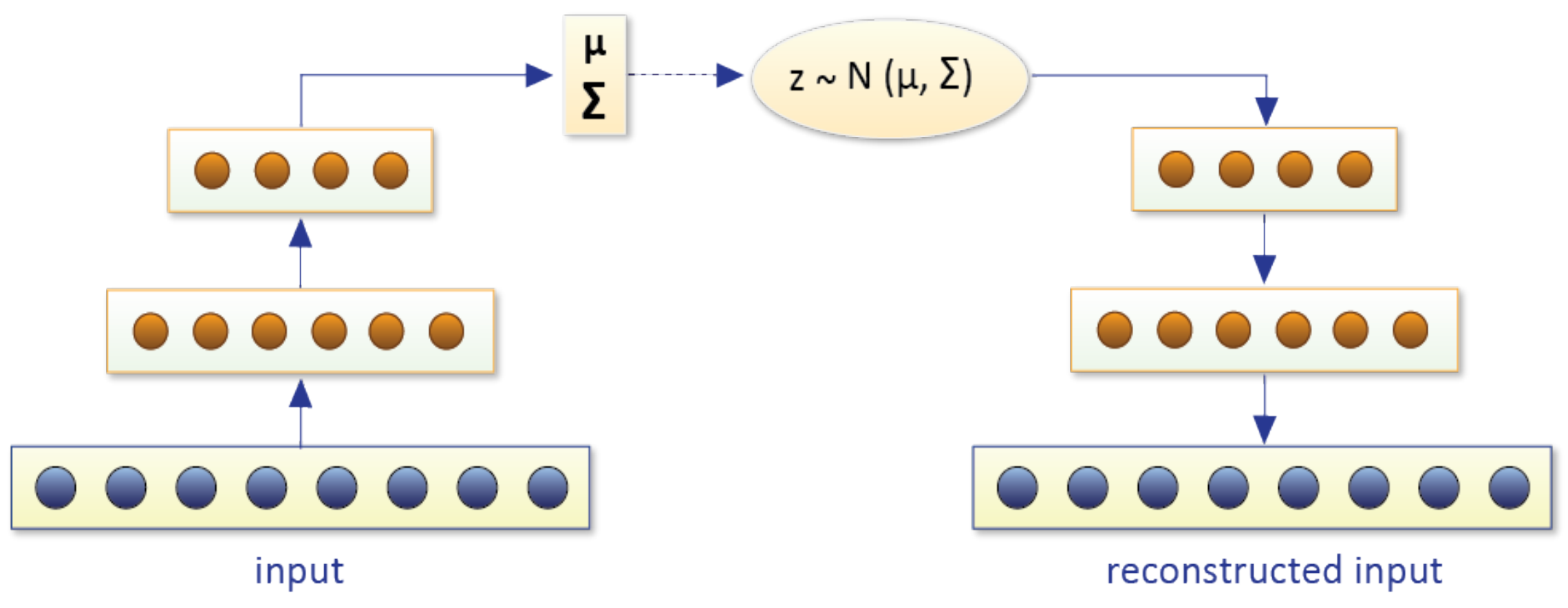
Formulation of the problem: Let $\mathbf{z}$ be a continuous latent random variable, generating the observation $\mathbf{x}$, through a function $f_\theta(\cdot)$, such that $\int p_\theta(\mathbf{x}|\mathbf{z})p_\theta(\mathbf{z})d\mathbf{z}$ is intractable. The goal is *inference*, i.e. $p_\theta(\mathbf{z}|\mathbf{x})$

Variational inference finds $q_\phi(\mathbf{z}|\mathbf{x}) \approx p_\theta(\mathbf{z}|\mathbf{x})$, by optimizing a lower-bound of the log-likelihood:

$$\ln p_\theta(\mathbf{x}) \geq \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\ln p_\theta(\mathbf{x}|\mathbf{z})] - \mathrm{KL}(q_\phi(\mathbf{z}|\mathbf{x}) \parallel p_\theta(\mathbf{z}))$$



Variational Autoencoders tackle the problem of *approximate inference* in the context of *neural networks*.



Limitation: assume independence of the latent variables and the observations. Recent papers, such as [5] and [1] propose methods of enriching the distribution of the latent variables.

Goal: introduce *correlations* between latent variables, through a Gaussian distribution with a sparse precision matrix, by using *Gaussian Graphical Models* (GGMs) [3].

Gaussian Graphical Models for Variational Autoencoders

Proposed Models: *chain* and *grid* topologies $\Longrightarrow$ sparse precision matrix $\mathbf{P}$; the number of non-zero elements of $\mathbf{P}$ is linear in the dimension of the latent variable. Similarly, sampling and computing the KL can be done in linear time.

Key observation: the neural network can learn *any ordering* for the latent variable. Thus, model selection is invariant to permutations of the nodes.

Our Approach:

- learn the values of the Cholesky decomposition $\mathbf{L}\mathbf{L}^\mathbf{T} = \mathbf{P}$ through the network
- sample in linear time by solving the linear system $\mathbf{L}^\mathbf{T}\mathbf{z} = \epsilon$, where $\epsilon \sim \mathcal{N}(0, I)$

Chain Dependency Structure

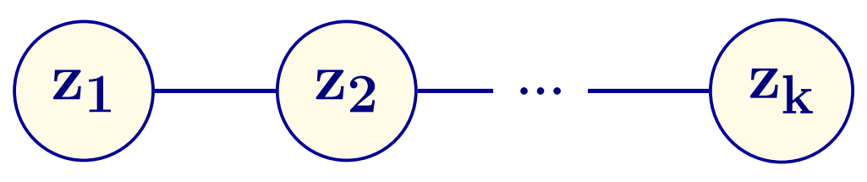


Figure 1: Chain Model

$$P = \begin{pmatrix} \sigma_1 & \lambda_1 & & 0 \\ \lambda_1 & \sigma_2 & \lambda_2 & \\ & & \ddots & \\ 0 & & \lambda_{k-1} & \sigma_k \end{pmatrix}$$

- The GGM in [1] is represented by a *tridiagonal* $\mathbf{P} = \Sigma^{-1}$ for $p(\mathbf{z}|\mathbf{x}) = \mathcal{N}(\mu, \Sigma)$;
- compute $\mathrm{Tr}(\Sigma)$ using the tridiagonal matrix inversion algorithm, in linear time.

Grid Dependency Structure

- For fig. [3] $\mathbf{P}$ is *block-tridiagonal* given by a *lower-block-bidiagonal Cholesky factor*;
- Approximate $\mathrm{Tr}(\Sigma)$ with stochastic methods or exact computation in quadratic time
- For fig. [2] we can guarantee that $\mathbf{P}$ is positive definite with Gershgorin circle theorem.

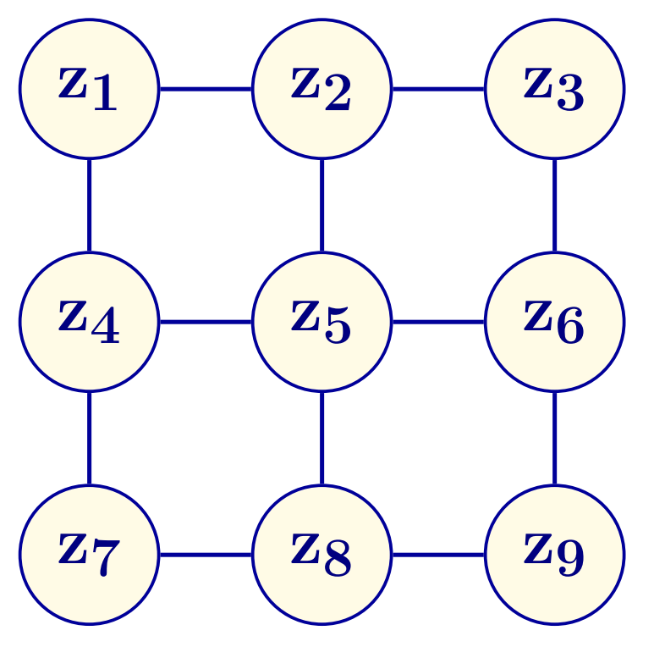


Figure 2: Regular Grid

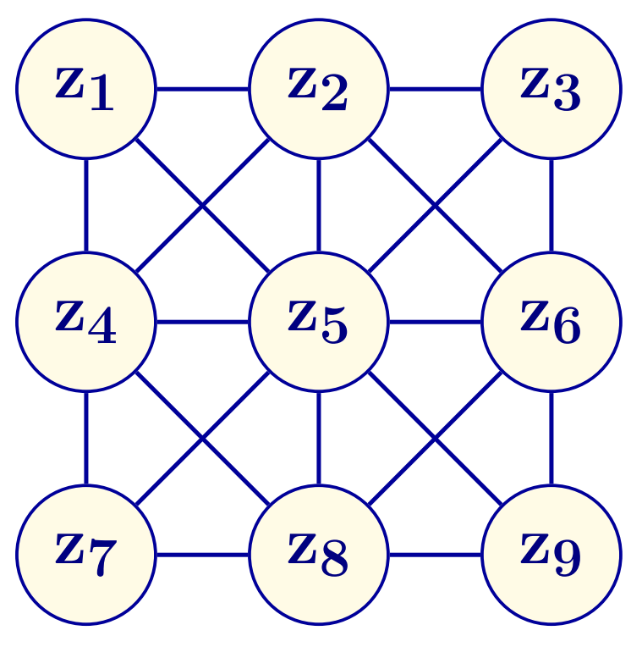


Figure 3: Extended Grid

Graphical Models for the Observed Variable

- A graphical model can be adapted also for the generative network
- For $\mathbf{x} \in (0,1)^n$ continuous, assume a multivariate logit-normal distribution:

$$p(x|z) = \frac{1}{\prod_{i=1}^n (x_i(1-x_i))}\mathcal{N}(\log \frac{x}{1-x} \mid \mu, \Sigma)$$

- The *grid* is a *natural model for images*, capturing *correlations between adjacent pixels*.

Experimental Results

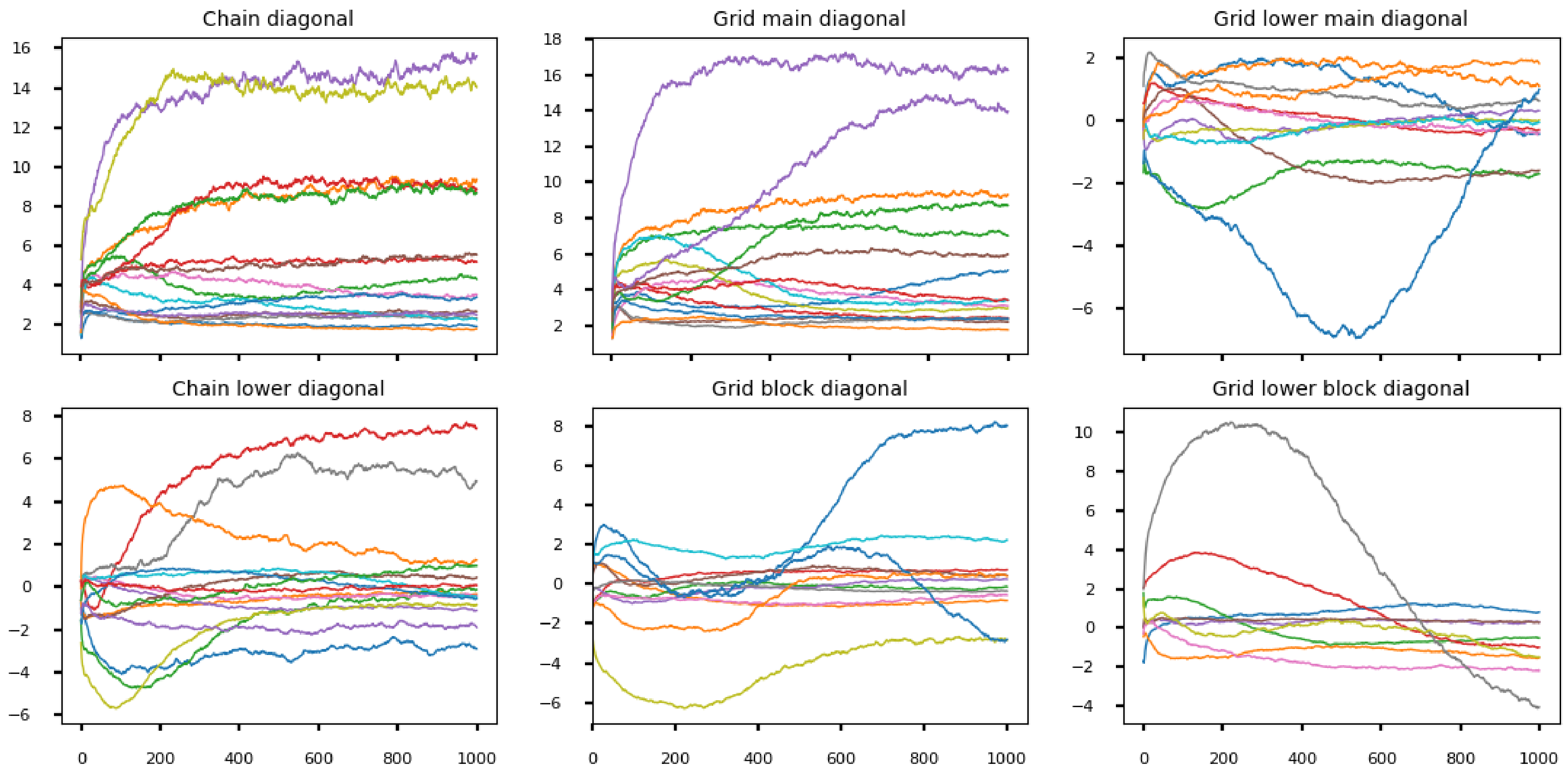


Figure 4: Diagonals of the Cholesky factor for the chain and grid for one data point

- Comparison on stochastically binarized MNIST between three different models for the latent variables;
- *network architecture*: feed-forward with two hidden layers, latent size 16, exponential linear units;
- plots for the means in fig. [5] and for the elements of the Cholesky factors in the grid and chain structures in fig. [4], by tracking one point in the dataset.

Model	$\approx \log p(x)$	$\log p(x) \geq$
Diagonal	-89.2	-93.4
Chain	-89.1	-93.4
Ext. Grid	-88.8	-93.0

Table 1: Approx. log-marginal and lower-bound

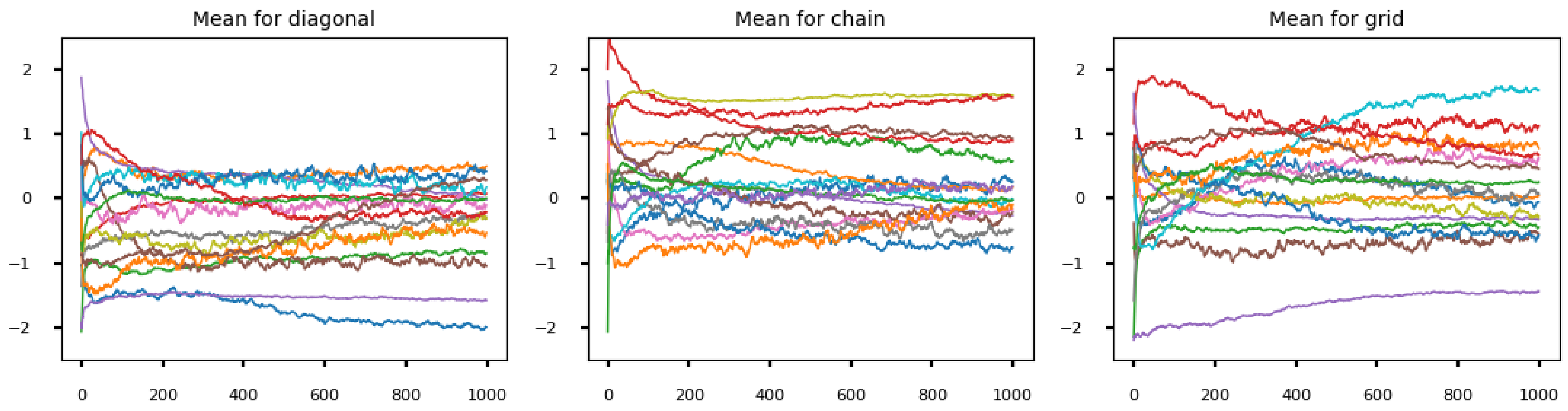


Figure 5: Means for the diagonal, chain and grid for one data point

Future Research Directions

- Experimental analysis of the impact on the log-likelihood when using the grid structures to introduce correlations between observations;
- test the impact of the Gaussian graphical models for the latent variables on multiple stochastic layers;
- explore the properties of the space of latent variables, using the geometry of the statistical models associated;

References

[1] Diederik P Kingma, Tim Salimans, Rafal Jozefowicz, Xi Chen, Ilya Sutskever, and Max Welling. Improving variational autoencoders with inverse autoregressive flow. In *Advances In Neural Information Processing Systems*, pages 4736–4744, 2016.

[2] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[3] Steffen L Lauritzen. *Graphical models*, volume 17. Clarendon Press, 1996.

[4] Alexandra Peşte and Luigi Malagó. Towards the use of gaussian graphical models in variational autoencoders. *ICML 2017 Workshop on Implicit Models*, 2017.

[5] Danilo Jimenez Rezende and Shakir Mohamed. Variational inference with normalizing flows. *arXiv preprint arXiv:1505.05770*, 2015.

[6] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pages 1278–1286, 2014.