Project REBAR

Maja Andrzejczuk, Piotr Bielecki, Maciej Orsłowski

Abstract

Learning in models with discrete latent variables is challenging due to high variance of gradient estimators. In this report, we aim to reproduce the **REBAR** method, introduced by Tucker et al. (2017). It combines the REINFORCE estimator with continuous relaxation of discrete variables, through a control variate, producing **unbiased**, **low-variance** gradient estimates.

1 Introduction

A stochastic gradient descent is a state-of-the art method for optimisation. However, for it to work properly, it requires an unbiased estimator for gradients (see Appendix C for SGD's proof of convergence). Not only that, but also SGD is only as good as small the variance of the estimator.

One solution for gradient estimation is to utilise the REINFORCE estimator. It is unbiased, however it has high variance. On the other hand, we could parametrize and apply softening, but the reparametrization trick estimator achieved in this way is biased.

This is where REBAR estimator comes useful. It combines REINFORCE method with continuous relaxation through a control variate, which results in reduced variance, while still being unbiased.

2 Preliminaries

2.1 Problem introduction

We consider a simple scenario. Let $b \sim \text{Bernoulli}(\theta)$ be a vector of independent binary random variables parametrized by θ. We wish to maximize

$$\mathop{\mathbb{E}}_{p_\theta(b)}[f(b,\theta)]$$

Where $f(b,\theta)$ is differentiable with respect to b and θ.

2.2 REINFORCE

Typically, this problem has been approached by gradient ascent, which requires efficiently estimating the REINFORCE estimator (details in appendix, section A), represented on the right-hand side of equation 1.

Following the original authors, in equations besides this, we suppress the dependence of f on θ, assuming $f(b,\theta) = f(b)$, and we concentrate on the second term, as stated in the paper:

(...)the dependence of f(b, θ) on θ is straightforward to account for(...)

$$\frac{\partial}{\partial\theta}\mathop{\mathbb{E}}_{p_\theta(b)}[f(b,\theta)] = \mathop{\mathbb{E}}_{p_\theta(b)}\left[\frac{\partial}{\partial\theta}f(b,\theta) + f(b,\theta)\frac{\partial}{\partial\theta}log\ p_\theta(b)\right] \tag{1}$$

*In practice, the first term can be estimated effectively with a single Monte Carlo sample, however a single sample estimator of the second term has **high variance**.*

2.3 Variance reduction through control variates

To counteract this, authors of the REBAR paper mention several other works showing, that a carefully designed control variate can reduce the variance of the second term on the right-hand side of equation 1 significantly. We can subtract any c (random or constant) as long as we can correct the bias (for details - see appendix, section B):

$$\frac{\partial}{\partial\theta}\mathop{\mathbb{E}}_{p_\theta(b,c)}[f(b)] = \frac{\partial}{\partial\theta}\left(\mathop{\mathbb{E}}_{p_\theta(b,c)}[f(b)-c] + \mathop{\mathbb{E}}_{p_\theta(b,c)}[c]\right) =$$
$$= \mathop{\mathbb{E}}_{p_\theta(b,c)}\left[(f(b)-c)\frac{\partial}{\partial\theta}\log p_\theta(b)\right] + \frac{\partial}{\partial\theta}\mathop{\mathbb{E}}_{p_\theta(b,c)}[c]$$

2.4 Reparametrization

Alternatively, we can parametrize (for details on reparametrization trick, see appendix, section D) b as $b = H(z)$, where H is an element-wise hard threshold function:

$$H(z) = \mathbb{1}[z \geq 0]$$

and z is a vector of independent random variables. In the REBAR paper, the authors propose the usage of logistic random variables defined by:

$$z := \log\frac{\theta}{1-\theta} + \log\frac{u}{1-u}$$

where $u \sim \text{Uniform}[0,1]$. We use an alternative:

$$z := Z(u,\theta) := \Phi^{-1}(u) + \theta$$

where $\Phi^{-1}(x)$ is the inverse cumulative distribution function (quantile function) of the standard normal distribution. Modelled in such a way, z is normally distributed: $z \sim N(\theta, 1)$.
However, the discontinuity of the hard threshold prevents us from using the reparametrization trick directly.

2.5 Continuous relaxation of discrete variables

Replacing all occurences of the hard threshold with a continuous relaxation using sigmoid function (depicted in figure 1):

$$H(z) \approx \sigma_\lambda(z) := \sigma\left(\frac{z}{\lambda}\right) = \left(1 + \exp\left(-\frac{z}{\lambda}\right)\right)^{-1}$$

Allows us to compute low-variance gradient estimates for the relaxed model that approximate the gradient for the discrete model. In summary:

$$\frac{\partial}{\partial\theta}\mathop{\mathbb{E}}_{p_\theta(b)}[f(b)] = \frac{\partial}{\partial\theta}\mathop{\mathbb{E}}_{p_\theta(z)}[f(H(z))] \approx \frac{\partial}{\partial\theta}\mathop{\mathbb{E}}_{p_\theta(z)}[f(\sigma_\lambda(z))] = \mathop{\mathbb{E}}_{p(u)}\left[\frac{\partial}{\partial\theta}f(\sigma_\lambda(Z(u,\theta)))\right] \qquad (2)$$

where λ is a temperature parameter, controlling the tightness of the relaxation.
This results in a low variance, but ***biased*** *Monte Carlo estimator for the discrete model. As $\lambda \to 0$ the approximation becomes exact, but the variance of the Monte Carlo estimator diverges. In practice, λ must be tuned to balance bias and variance.*

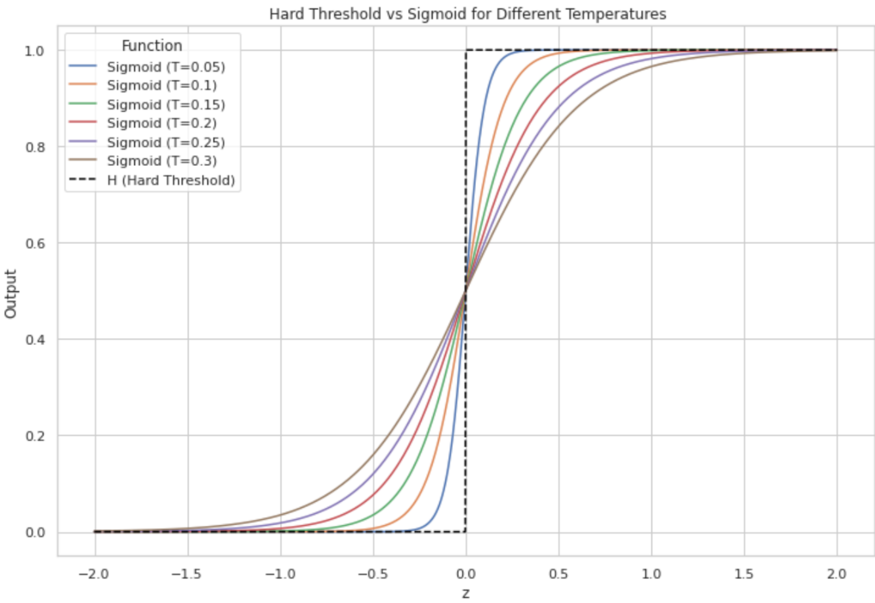


Figure 1: Analyzing Sigmoid Function Behavior under Temperature Variations and Hard Threshold

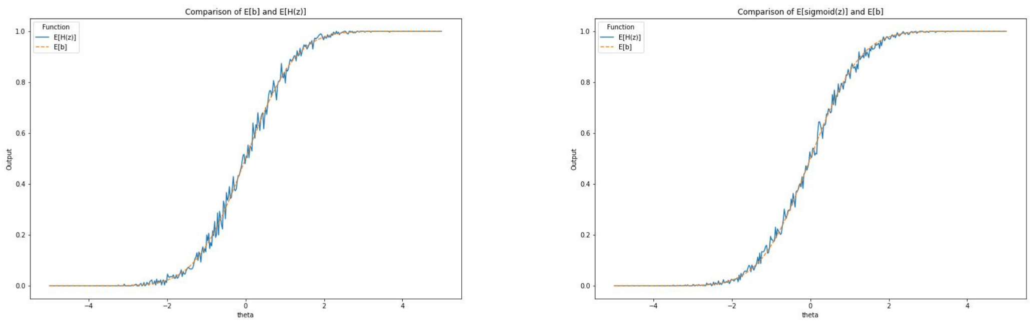


Figure 2: A comparative analysis of $\mathbb{E}[b]$ (dashed orange line) with $\mathbb{E}[H(z)]$ and $\mathbb{E}[\text{sigmoid}(z)]$ respectively (blue line). The results indicate that the application of the sigmoid function is unnecessary for estimating the expected value, as softening does not provide additional benefits in this context.

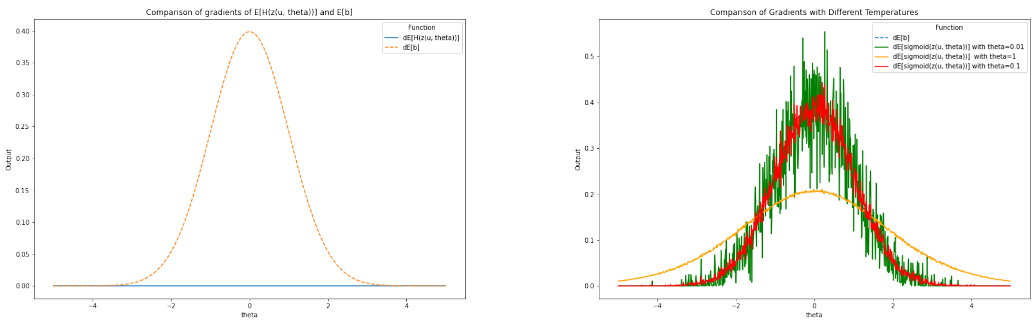


Figure 3: The left figure shows the gradient comparison between the expected values of $H(z)$ and b, indicating that softening is necessary as it allows the gradients to be approximated. The right figure demonstrates how different temperatures impact the gradients of $\mathbb{E}[\text{sigmoid}(z)]$, with higher temperatures leading to sharper, but more biased approximations.

3 The REBAR estimator

To reiterate - we seek a low-variance, unbiased gradient estimator. Inspired by the Concrete relaxation, the original authors aim to construct a control variate based on the difference between the REINFORCE gradient estimator for the relaxed model and the gradient estimator from the reparametrization trick.
The form of the REINFORCE gradient estimator for the relaxed model is:

$$\frac{\partial}{\partial\theta}\underset{p_\theta(z)}{\mathbf{E}}[f(\sigma_\lambda(z)] = \underset{p_\theta(z)}{\mathbf{E}}\left[f(\sigma_\lambda(z)\frac{\partial}{\partial\theta}\log p(z)\right]$$

Authors of the REBAR paper state, that their key insight is a conditional marginalization for the control variate $(f(\sigma_\lambda(z)))$:

$$\underset{p_\theta(z)}{\mathbf{E}}\left[f(\sigma_\lambda(z))\frac{\partial}{\partial\theta}\log p(z)\right] = \underset{p_\theta(b)}{\mathbf{E}}\left[\frac{\partial}{\partial\theta}\underset{p_\theta(z|b)}{\mathbf{E}}[f(\sigma_\lambda(z))]\right] + \underset{p_\theta(b)}{\mathbf{E}}\left[\underset{p_\theta(z|b)}{\mathbf{E}}[f\sigma_\lambda(z))]\frac{\partial}{\partial\theta}\log p(b)\right] \quad (3)$$

The first term on the right-hand side can be estimated using the reparametrization trick:

$$\underset{p_\theta(b)}{\mathbf{E}}\left[\frac{\partial}{\partial\theta}\underset{p_\theta(z|b)}{\mathbf{E}}[f(\sigma_\lambda(z))]\right] = \underset{p_\theta(b)}{\mathbf{E}}\left[\underset{p(v)}{\mathbf{E}}\left[\frac{\partial}{\partial\theta}f(\sigma_\lambda(\tilde{z}))\right]\right]$$

Where $v \sim \text{Uniform}[0,1]$ and $\tilde{z} = \tilde{z}(v, H(z), \theta)$ is the reparametrization for $z|b$. In our case, this is:

$$\tilde{z}(v, H(z), \theta) = \begin{cases} \theta + \Phi^{-1}(\Phi(\theta)v) & , H(z) = 0 \\ \theta - \Phi^{-1}(\Phi(-\theta)v) & , H(z) = 1 \end{cases}$$

Depicted in Figure 4, is a plot showing the distribution of z, along with the conditional distributions of $\tilde{z}$ for $\theta = 1$. Both the conditional distributions resemble that of z's - when the cdf of z is split along $y = 0$, the conditional cdfs are "vertically scaled versions" of one of the splits.

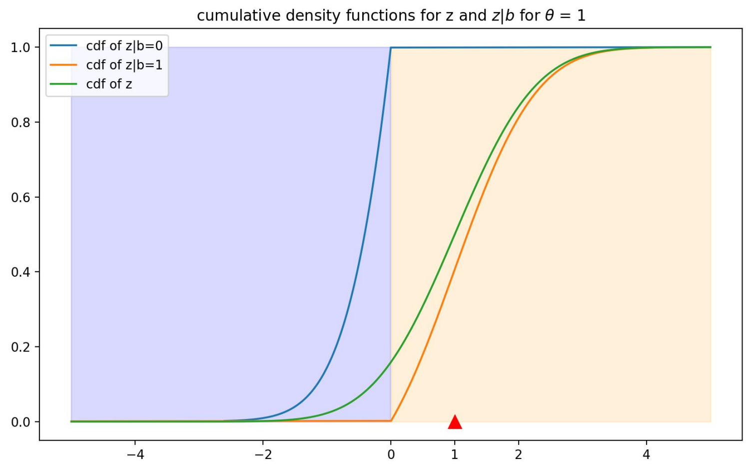


Figure 4: Comparison of the densities of $\tilde{z}$ under different values of $H(z)$

Applying the reparametrization to the control variate's REINFORCE gradient estimator (equation 3) yields:

$$\underset{p_\theta(z)}{\mathbf{E}}\left[f(\sigma_\lambda(z))\frac{\partial}{\partial\theta}\log p(z)\right] = \underset{p_\theta(b)}{\mathbf{E}}\left[\underset{p_\theta(v)}{\mathbf{E}}\left[\frac{\partial}{\partial\theta}f(\sigma_\lambda(\tilde{z}))\right]\right] + \underset{p_\theta(b)}{\mathbf{E}}\left[\underset{p_\theta(z|b)}{\mathbf{E}}[f\sigma_\lambda(z))]\frac{\partial}{\partial\theta}\log p_\theta(b)\right]$$

Combining the REINFORCE estimator with a control variate, which we correct with the reparametrization trick gradient, we arrive at the REBAR estimator:

$$\frac{\partial}{\partial \theta} \underset{p_\theta(b)}{\mathbb{E}}[f(b)] = \underset{p(u,v)}{\mathbb{E}} \left[\left[f(H(z)) - \eta f(\sigma_\lambda(\tilde{z})) \right] \frac{\partial}{\partial \theta} \log p_\theta(b) \bigg|_{b=H(z)} + \eta \frac{\partial}{\partial \theta} f(\sigma_\lambda(z)) - \eta \frac{\partial}{\partial \theta} f(\sigma_\lambda(\tilde{z})) \right] \tag{4}$$

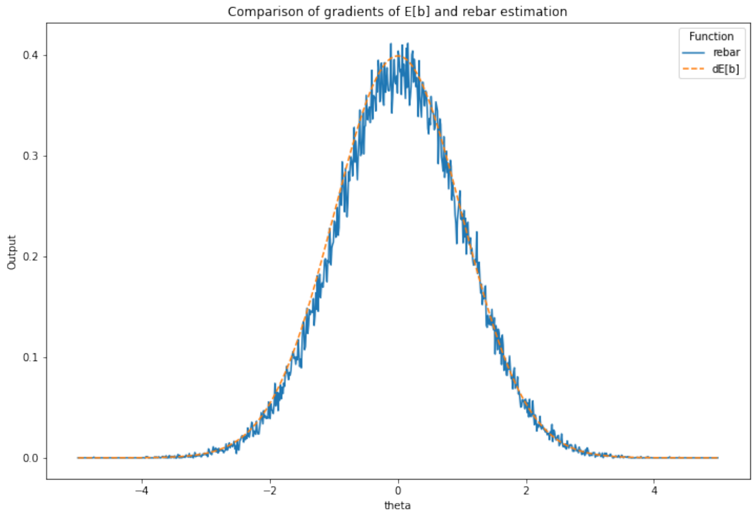


Figure 5: Comparison of gradients of E[b] and REBAR estimation

4 Gaussian Mixture Model

In our experiments, we use REBAR to estimate gradients for Stochastic Gradient Descent aimed to optimise parameters in the Gaussian Mixture Model (GMM). GMM is a model where observations are generated from a combination of multiple normal distributions. Each of these distributions is a component (or class) of GMM, and they have a weight specified, which is a probability that an observation comes from that component. GMM with k components can be described by the following formula:

$$X(\zeta, j, \mu, \Sigma) = \Sigma_j \cdot \zeta + \mu_j$$

where $j = 1, ..., k$ specifies a component ($\sum_{j=1}^{k} p(c = j) = 1$, and $\zeta \sim N(0, 1)$. In the case of our experiments, we exclusively use a Gaussian Mixture Model with two components. In this scenario, the only weight that needs to be specified is $\mathbb{P}(c = 1)$, which we model through a parameter θ in the following way, applying parametrization: $C = C(z) = \mathbb{1}[Z \geq 0]$ where $Z \sim \mathbf{N}(\theta, \mathbf{1})$, which gives us:

$$\mathbb{P}(c = 1) = \Phi(-\theta)$$

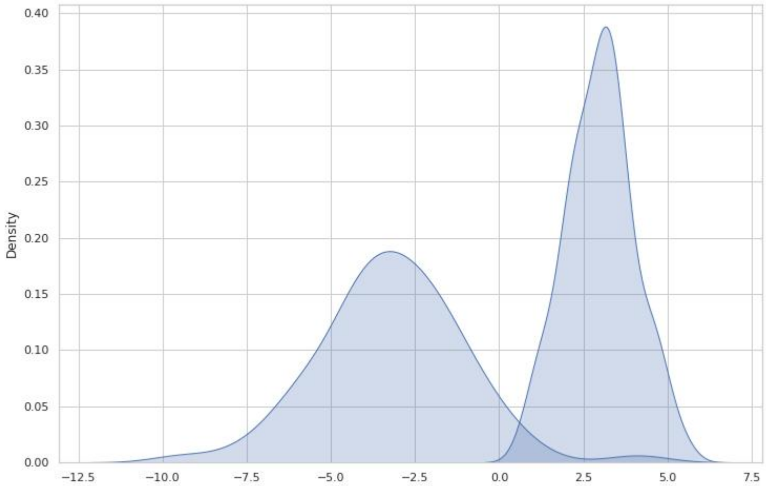


Figure 6: Example of a two-component Gaussian Mixture Model (GMM) for 1D data. The plot shows kernel density estimations for a sample of two Gaussian distributions with different means and covariance matrices.

In the context of REBAR, c is a discrete latent variable in our model. The function we wish to maximise is a likelihood that an observed sample X comes from the Gaussian Mixture Model. This function we define as:

$$p(X|c, \mu, \Sigma) = c \cdot \phi_{\mu_0,\Sigma_0}(X) + (1-c) \cdot \phi_{\mu_1,\Sigma_1}(X)$$

where c is the weight of class 1, and ϕ_{μ_j,Σ_j} is a density function of a multivariate normal distribution with means μ_j and covariance matrix Σ_j. In practice, maximising the likelihood was not feasible, as:

$$\frac{\partial}{\partial\theta}\mathbb{L}(\theta, \mu, \Sigma) = \frac{\partial}{\partial\theta}\prod_{i=1}^{n} p(x_i|\theta, \mu, \sigma) =$$

$$= \frac{\partial}{\partial\theta}\prod_{i=1}^{n}\left[\sum_{j=1}^{\#\ classes} p(x_i|\theta, \mu, \Sigma)p(c = j|\theta)\right] = \frac{\partial}{\partial\theta}\prod_{i=1}^{n}\left[\mathop{\mathbb{E}}_{p_\theta(c)}(p(x_i|c, \mu, \Sigma))\right]$$

Therefore, we aim to maximize log-likelihood:

$$\frac{\partial}{\partial\theta}\mathbf{l}(\theta, \mu, \Sigma) = \frac{\partial}{\partial\theta}\sum_{i=1}^{n}\log p(x_i|\theta, \mu, \Sigma) = \sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log p(x_i|\theta, \mu, \Sigma)$$

$$= \sum_{i=1}^{n}\left[\frac{\frac{\partial}{\partial\theta}p(x_i|\theta, \mu, \Sigma)}{p(x_i|\theta, \mu, \Sigma)}\right] = \sum_{i=1}^{n}\left[\frac{\frac{\partial}{\partial\theta}\mathbb{E}_{p_\theta(c)}(p(x_i|c, \mu, \Sigma))}{\mathbb{E}_{p_\theta(c)}(p(x_i|c, \mu, \Sigma))}\right]$$

Where we apply REBAR to the nominator of equation above. The denominator is approximated well with a monte-carlo sample.

We estimate gradients with respect to θ using REBAR, while for μ_j and Σ_j we calculate the gradient normally(details in E.5). The REBAR gradient estimation for θ in our case can be formulated as:

$$\frac{\partial}{\partial\theta}\mathbb{E}[p(X|\theta, \mu, \Sigma)] = \mathop{\mathbb{E}}_{p(u,v)}\left[\left[p(X_i|H(z), \mu, \Sigma) - p(X_i|\sigma_\lambda(\tilde{z}), \mu, \Sigma)\right]\frac{\partial}{\partial\theta}\log p(c|\theta)\Big|_{c=H(z)}\right. \tag{5}$$
$$\left. + \frac{\partial}{\partial\theta}p(X_i|\sigma_\lambda(z), \mu, \Sigma) - \frac{\partial}{\partial\theta}p(X_i|\sigma_\lambda(\tilde{z}), \mu, \Sigma)\right]$$

In Figure 7, we present a graphical model for our implementation of REBAR.

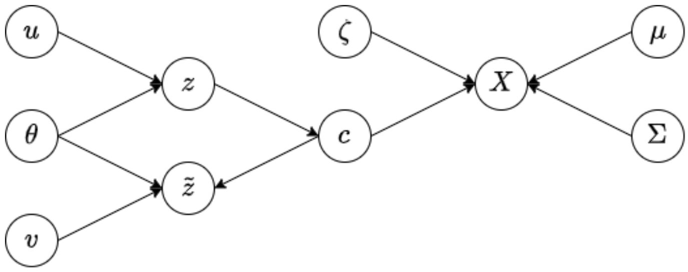
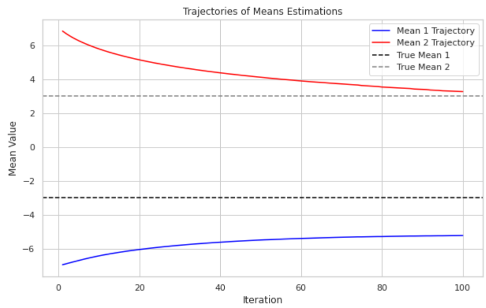


Figure 7: A graphical model for REBAR with GMM

4.1 Results

In this section, we present some of our results. In figures 8 and 9 trajectories for a 1D and 2D model are depicted. In figure 10 a comparison of the default EM algorithm used for GMM models,



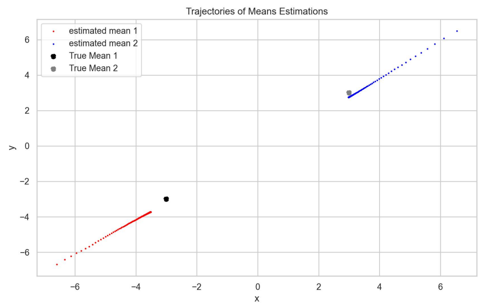


Figure 8: Trajectories of means estimations - 1D; Figure 9: Trajectories of means estimations - 2D

our implementation of REBAR estimator, and reinforce estimator for the relaxed model(details in E.6).

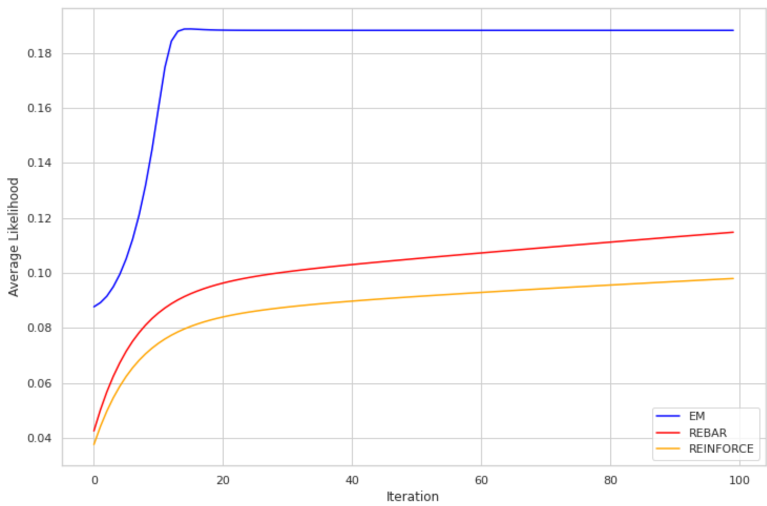


Figure 10: Average Likelihood Over Iterations for EM, REBAR, and REINFORCE

Appendix

A REINFORCE estimator

The REINFORCE estimator utilizes a differentiation rule called the log-derivative trick:

$$\nabla_\theta p_\theta(x) = p_\theta(x) \nabla_\theta \log p_\theta(x) \tag{6}$$

which is an application of the differentiation rule to the logarithm. With this, estimation of the gradient can be rewritten as follows:

$$\begin{aligned}
\nabla_\theta \mathop{\mathbb{E}}_{x \sim p_\theta(x)} [f(x)] &= \nabla_\theta \int f(x) p_\theta(x) dx \\
&= \int f(x) \nabla_\theta p_\theta(x) dx \\
&= \int f(x) p_\theta(x) \nabla_\theta \log p_\theta(x) dx \\
&= \mathop{\mathbb{E}}_{x \sim p_\theta(x)} [f(x) \nabla_\theta \log p_\theta(x)]
\end{aligned}$$

It's also interesting to note that REINFORCE places no restriction on the nature of the function f, and it does not even need to be differentiable for us to estimate the gradients of its expected value.

B Variance reduction through control variates

We want to estimate $\mathbb{E}_x[f(x)]$ for an arbitrary function f. The variance of the naïve Monte Carlo estimator $\mathbb{E}_x[f(x)] \approx \frac{1}{n} \sum_i f(x_i)$, where $x_1, ..., x_n \sim p(x)$ can be reduced by introducing a control variate $g(x)$:

$$\mathbb{E}[f(x)] \approx \left(\frac{1}{k} \sum_i f(x_i) - \eta g(x_i) \right) + \eta \, \mathbb{E}[g(x)]$$

The above is an unbiased estimator for any value of η. To minimize the variance of the estimator, the optimal choice for η is

$$\eta = \frac{\mathrm{Cov}(f, g)}{\mathrm{Var}(g)}$$

and it reduces the variance of the estimator by $1 - \rho(f, g)^2$, where $\rho = Corr(f, g)$. If we cannot compute $\mathbb{E}[g]$, we can use a low-variance estimator $\hat{g}$.

B.1 Underlying principle

Let the unknown parameter of interest be μ, and we assume we have a statistic m such that $\mathbb{E}(m) = \mu$, meaning m is an unbiased estimator for μ. Suppose we calculate another statistic t such that $\mathbb{E}(t) = \tau$ is a known value. Then:

$$m^* = m + \eta(t - \tau)$$

is an unbiased estimator for μ for any choice of the coefficient η. The variance of the resulting m^* estimator is:

$$\mathrm{Var}(m^*) = \mathrm{Var}(m) + \eta^2 \mathrm{Var}(t) + 2\eta \mathrm{Cov}(m, t)$$

the optimal value for η is calculated by differentiating the above with respect to η.

C SGD - proof of convergence

In this section of the appendix, we provide a proof of convergence for the Stochastic Gradient Descent for convex and smooth functions, as in Garrigos & Gower (2024).

This problem can be formulated as follows. Let $(x_t)_{t\in\mathbb{N}}$ be a sequence generated by the SGD algorithm, and $(\gamma_t)_{t\in\mathbb{N}}$ a sequence of step sizes satisfying $0 < \gamma_t < \frac{1}{4L_{\max}}$. We aim to prove that for every $T \geq 1$:

$$\mathbb{E}\big[f(\bar{x}_T) - \inf f\big] \leq \frac{\|x_0 - x^*\|^2}{\sum_{t=0}^{T-1}\gamma_t} + 2\sigma_f^* \frac{\sum_{t=0}^{T-1}\gamma_t^2}{\sum_{t=0}^{T-1}\gamma_t}$$

where $\bar{x}_T := \left(\sum_{t=0}^{T-1}\gamma_t\right)^{-1} \sum_{t=0}^{T-1} x_t$.

The proof proceeds as follows. Let $x^* \in \operatorname{argmin} f$, so we have $\sigma_f^* = \mathbb{V}[\nabla f_i(x^*)]$. Moving forward, $\mathbb{E}_t[\cdot|x_t]$ will be denoted by $\mathbb{E}_t[\cdot]$ to simplify notations. Let us begin by analysing the following expression:

$$\|x_{t+1} - x^*\|^2 = \|x_t - x^*\|^2 - 2\gamma_t\langle\nabla \underline{f_i}(x_t), x_t - x^*\rangle + \gamma_t^2\|\nabla f_i(x_t)\|^2$$

Hence, after taking the expectation conditioned on x_t, we can use the convexity of f and a variance transfer lemma to write:

$$\begin{aligned}
\mathbb{E}_t\left[\|x_{t+1} - x^*\|^2\right] &= \|x_t - x^*\|^2 + 2\gamma_t\langle\nabla \underline{f}(x_t), x^* - x_t\rangle + \gamma_t^2\underline{\mathbb{E}_t[\|\nabla f_i(x_t)\|^2]} \leq \\
&\leq \|x_t - x^*\|^2 + 2\gamma_t(2\gamma_t L_{\max} - 1)(f(x_t) - \inf f) + 2\gamma_t^2\sigma_f^* \leq \\
&\leq \|x^* - x_t\|^2 - \gamma_t(f(x_t) - \inf f) + 2\gamma_t^2\sigma_f^*
\end{aligned}$$

The last inequality uses the assumption of $\gamma_t L_{\max} \leq \frac{1}{4}$. We can rearrange this and take the expected value to get:

$$\gamma_t\mathbb{E}\big[(f(x_t) - \inf f)\big] \leq \mathbb{E}\left[\|x^* - x_t\|^2\right] - \mathbb{E}\left[\|x_{t+1} - x^*\|^2\right] + 2\gamma_t^2\sigma_f^*$$

Then, by summing over $t = 0, 1, ..., T-1$ for $T \geq 1$ and using telescopic cancellation, we obtain:

$$\sum_{t=0}^{T-1}\gamma_t\mathbb{E}\big[f(x_t) - \inf f\big] \leq \|x_0 - x^*\|^2 - \mathbb{E}\left[\|x_T - x^*\|^2\right] + 2\sigma_f^*\sum_{t=0}^{T-1}\gamma_t^2$$

Since $\mathbb{E}\left[\|x_T - x^*\|^2\right] \geq 0$, dividing both sides by $\sum_{t=0}^{T-1}\gamma_t$ gives:

$$\mathbb{E}\left[\sum_{t=0}^{T-1}\frac{\gamma_t}{\sum_{t=0}^{T-1}\gamma_t}(f(x_t) - \inf f)\right] \leq \frac{\|x_0 - x^*\|^2}{\sum_{t=0}^{T-1}\gamma_t} + 2\sigma_f^*\frac{\sum_{t=0}^{T-1}\gamma_t^2}{\sum_{t=0}^{T-1}\gamma_t}$$

Finally, by using both the fact that f is convex and Jensen's inequality, we arrive at:

$$\mathbb{E}\big[f(\bar{x}_T) - \inf f\big] \leq \mathbb{E}\left[\sum_{t=0}^{T-1}\frac{\gamma_t}{\sum_{t=0}^{T-1}\gamma_t}(f(x_t) - \inf f)\right] \leq \frac{\|x_0 - x^*\|^2}{\sum_{t=0}^{T-1}\gamma_t} + 2\sigma_f^*\frac{\sum_{t=0}^{T-1}\gamma_t^2}{\sum_{t=0}^{T-1}\gamma_t}$$

which ends the proof.

The proof shows the need for an unbiased estimator (the transition underscored in red holds only for an unbiased gradient estimator), as well as the relation between the variance of the estimator and speed of convergence of SGD (underscored in blue; the relation is linear).

D Reparametrization Trick

(...)Recall that the problem in evaluating $\frac{\partial}{\partial\theta}\mathbf{E}_{p_\theta(b)}[f(b)]$ is the fact that the expectation is taken with respect to a distribution with parameters θ and we can't compute the derivative of that stochastic quantity. Reparametrization gradients (also known as pathwise gradients) allow us to compute this by re-writing the samples of the distribution p_θ in terms of a noisy variable ζ, independent of θ. - Javed (2018)

$$\begin{aligned}
\zeta &\sim q(\zeta) \\
x &= g_\theta(\zeta) \\
\nabla_\theta\mathbb{E}_{x\sim p_\theta(x)}[f(x)] &= \nabla_\theta\mathbb{E}_{\zeta\sim q(\zeta)}[f(g_\theta(\zeta))] \\
&= \mathbb{E}_{\zeta\sim q(\zeta)}[\nabla_\theta f(g_\theta(\zeta))]
\end{aligned}$$

Where the last expected value is the reparametrization trick gradient estimator.

E Equation derivations

E.1 Equation 1

$$
\begin{aligned}
\frac{\partial}{\partial\theta}\mathop{\mathbb{E}}_{p_\theta(b)}[f(b)] = \frac{\partial}{\partial\theta}\int f(b)p_\theta(b)db = \int \frac{\partial}{\partial\theta}[f(b)p_\theta(b)]db = \\
= \int\left[\frac{\partial}{\partial\theta}f(b)p_\theta(b) + f(b)\frac{\partial}{\partial\theta}p_\theta(b)\right]db = \int\left[\frac{\partial}{\partial\theta}f(b)p_\theta(b) + f(b)p_\theta(b)\frac{\partial}{\partial\theta}\log p_\theta(b)\right]db = \\
= \mathop{\mathbb{E}}_{p_\theta(b)}\left[\frac{\partial}{\partial\theta}f(b) + f(b)\frac{\partial}{\partial\theta}\log p_\theta(b)\right]
\end{aligned}
$$

Where the last equation in the first line is an application of Leibniz integral rule, while the last equation utilizes the log-derivative trick.

E.2 Equation 2

This is an application of the reparametrization trick with relaxation, allowing us to approximate the gradient by sampling.

E.3 Equation 3

Because $p(b|z) = 1 \implies p(z|b) = \frac{p(b|z)p(z)}{p(b)} = \frac{p(z)}{p(b)}$, we have $f(\sigma(z))\frac{\partial}{\partial\theta}\log p(z) = f(\sigma(z))\frac{\partial}{\partial\theta}(\log p(z|b) + \log p(b))$, which in turn yields equation 3:

$$
\begin{aligned}
\mathop{\mathbb{E}}_{p_\theta(z)}\left[f(\sigma(z))\frac{\partial}{\partial\theta}\log p(z)\right] = \mathop{\mathbb{E}}_{p_\theta(b)}\left[\mathop{\mathbb{E}}_{p(z|b)}\left[f(\sigma(z))\frac{\partial}{\partial\theta}\Big(\log p(z|b) + \log p(b)\Big)\right]\right] = \\
= \mathop{\mathbb{E}}_{p_\theta(b)}\left[\mathop{\mathbb{E}}_{p_\theta(z|b)}\left[f(\sigma(z))\frac{\partial}{\partial\theta}\log p(z|b)\right]\right] + \mathop{\mathbb{E}}_{p_\theta(b)}\left[\mathop{\mathbb{E}}_{p_\theta(z|b)}\left[f(\sigma(z))\frac{\partial}{\partial\theta}\log p(b)\right]\right] = \\
= \mathop{\mathbb{E}}_{p_\theta(b)}\left[\frac{\partial}{\partial\theta}\mathop{\mathbb{E}}_{p_\theta(z|b)}\left[f(\sigma(z))\right]\right] + \mathop{\mathbb{E}}_{p_\theta(b)}\left[\mathop{\mathbb{E}}_{p_\theta(z|b)}\left[f(\sigma(z))\right]\frac{\partial}{\partial\theta}\log p(b)\right]
\end{aligned}
$$

The last equation is an application of log-derivative trick for the first term, while for the second term we take $\frac{\partial}{\partial\theta}\log p(b)$ outside the expectation over $p_\theta(z|b)$, as it's not dependent on the distribution $p_\theta(z|b)$

E.4 Equation 4

$$
\frac{\partial}{\partial\theta}\mathop{\mathbb{E}}_{p(b)}[f(b)] = \frac{\partial}{\partial\theta}\mathop{\mathbb{E}}_{p(z)}[f(H(z)) - f(\sigma(z))] + \frac{\partial}{\partial\theta}\mathop{\mathbb{E}}_{p(z)}[f(\sigma(z))] \tag{7}
$$

We use REINFORCE on the first addend:

$$
\begin{aligned}
\frac{\partial}{\partial\theta}\mathop{\mathbb{E}}_{p(z)}[f(H(z)) - f(\sigma(z))] = \mathop{\mathbb{E}}_{p(z)}\left[[f(H(z)) - f(\sigma(z))]\frac{\partial}{\partial\theta}\log p(H(z))\right] = \\
= \mathop{\mathbb{E}}_{p(z)}\left[f(H(z))\frac{\partial}{\partial\theta}\log p(H(z))\right] - \mathop{\mathbb{E}}_{p(z)}\left[f(\sigma(z))\frac{\partial}{\partial\theta}\log p(H(z))\right] = \\
= \mathop{\mathbb{E}}_{p(b)}\left[f(b)\frac{\partial}{\partial\theta}\log p(b)\right] - \mathop{\mathbb{E}}_{p(b)}\left[\frac{\partial}{\partial\theta}\mathop{\mathbb{E}}_{p(z|b)}[f(\sigma(z))]\right] - \mathop{\mathbb{E}}_{p(b)}\left[\mathop{\mathbb{E}}_{p(z|b)}[f(\sigma(z))]\frac{\partial}{\partial\theta}\log p(b)\right] = \\
= \mathop{\mathbb{E}}_{p(b)}\left[\left(f(b) - \mathop{\mathbb{E}}_{p(z|b)}[f(\sigma(z))]\right)\frac{\partial}{\partial\theta}\log p(b)\right] - \mathop{\mathbb{E}}_{p(b)}\left[\mathop{\mathbb{E}}_{p(v)}\left[\frac{\partial}{\partial\theta}f(\sigma(\tilde{z}))\right]\right]
\end{aligned}
$$

Then going back to equation 7, we can write that:

$$\frac{\partial}{\partial\theta}\mathop{\mathbb{E}}_{p(b)}[f(b)] =$$

$$= \frac{\partial}{\partial\theta}\mathop{\mathbb{E}}_{p(z)}[f(\sigma(z))] + \mathop{\mathbb{E}}_{p(b)}\left[\left(f(b) - \mathop{\mathbb{E}}_{p(z|b)}[f(\sigma(z))]\right)\frac{\partial}{\partial\theta}\log p(b)\right] - \mathop{\mathbb{E}}_{p(b)}\left[\mathop{\mathbb{E}}_{p(v)}\left[\frac{\partial}{\partial\theta}f(\sigma(\widetilde{z}))\right]\right] =$$

$$= \frac{\partial}{\partial\theta}\mathop{\mathbb{E}}_{u,v}\left[f\big(\sigma(z(u,\theta))\big)\right] + \mathop{\mathbb{E}}_{u,v}\left[[f(H(z)) - f(\sigma(\widetilde{z}))]\frac{\partial}{\partial\theta}\log p(b)\bigg|_{c=H(z)}\right] - \mathop{\mathbb{E}}_{u,v}\left[\frac{\partial}{\partial\theta}f(\sigma(\widetilde{z}))\right] =$$

$$= \mathop{\mathbb{E}}_{u,v}\left[[f(H(z)) - f(\sigma(\widetilde{z}))]\frac{\partial}{\partial\theta}\log p(b)\bigg|_{c=H(z)} - \frac{\partial}{\partial\theta}f(\sigma(\widetilde{z})) + \frac{\partial}{\partial\theta}f\big(\sigma(z)\big)\right]$$

Where z and $\widetilde{z}$ are dependent only on the parameter of interest, and the random samples u and v as:

$$z = z(u,\theta) \quad \text{and} \quad \widetilde{z} = \widetilde{z}(v, H(z), \theta)$$
$$\text{for } H(z) = \mathbf{1}[z \geq 0]$$

E.5 Derivation of gradient for μ and Σ

We show the derivations for μ, Σ behaves exactly the same. Following the log-likelihood derivation from section about GMM:

$$\frac{\partial}{\partial\mu}\mathbf{l}(\theta,\mu,\Sigma) = \sum_{i=1}^{n}\frac{\frac{\partial}{\partial\mu}\mathbb{E}_{p_\theta(c)}(p(x_i|c,\mu,\Sigma))}{\mathbb{E}_{p_\theta(c)}(p(x_i|c,\mu,\Sigma))} = \sum_{i=1}^{n}\frac{\mathbb{E}_{p_\theta(c)}\frac{\partial}{\partial\mu}(p(x_i|c,\mu,\Sigma))}{\mathbb{E}_{p_\theta(c)}\,p(x_i|c,\mu,\Sigma)}$$

Which we could not do while calculating the derivative with respect to θ, because of the hard-threshold within $p_\theta(c)$

E.6 REINFORCE estimator for relaxed model

Here, we present the form of the estimator for the relaxed model used for comparison.

$$\frac{\partial}{\partial\theta}\log\mathbf{L}(\theta,\mu,\Sigma) = \sum_{i=1}^{n}\frac{\frac{\partial}{\partial\theta}\mathbb{E}_{p_\theta(c)}\,p(x_i|c,\mu,\Sigma)}{\mathbb{E}_{p_\theta(c)}\,p(x_i|c,\mu,\Sigma)} \approx \sum_{i=1}^{n}\frac{\mathbb{E}_{p_\theta(z)}\,p(x_i|\sigma_\lambda(z),\mu,\Sigma)\frac{\partial}{\partial\theta}\log p_\theta(z)}{\mathbb{E}_{p_\theta(z)}\,p(x_i|\sigma_\lambda(z),\mu,\Sigma)}$$

Where the nominator is the REINFORCE estimator for the relaxed model.

References

Garrigos, G. & Gower, R. M. (2024), 'Handbook of convergence theorems for (stochastic) gradient methods'.

Javed, S. A. (2018), 'REINFORCE vs Reparameterization Trick'.
URL: *https://stillbreeze.github.io/REINFORCE-vs-Reparameterization-trick/?fbclid=IwAR1il3HYJx4wAAqak4Wjy9eWxsB8qLK$_Y$bsWgiJ2OMb27nViqMPl8ILw − xU*

Tucker, G., Mnih, A., Maddison, C. J., Lawson, D. & Sohl-Dickstein, J. (2017), 'Rebar: Low-variance, unbiased gradient estimates for discrete latent variable models'.