1 Problem Definition

We consider loss functions \( J : \mathbb{R}^d \rightarrow \mathbb{R} \) whose gradient is implicitly defined as the expectation of a smooth vector-valued function \( Y : \mathbb{R}^n \times \mathbb{R}^d \rightarrow \mathbb{R}^d \).

The smooth function \( Y_a(\varepsilon, \theta) \) depends on a noise variable \( \varepsilon \in \mathbb{R}^n \) and on the parameters of interest \( \theta \in \mathbb{R}^d \). The noise variable \( \varepsilon \) is distributed according to an arbitrary density \( p(\varepsilon) \). More specifically, we are interested in gradients of the form

\[
\nabla_{\theta_a} J(\theta) = \mathbb{E}_{\varepsilon \sim p}[Y_a(\varepsilon, \theta)]
\]

where \( a = 1 \ldots d \).

A control variate is any function \( \Delta_a(\varepsilon, \theta) \) of the noise variables with known expectation (assumed here to be zero, i.e. \( \mathbb{E}_{\varepsilon \sim p}[\Delta_a(\varepsilon, \theta)] = 0 \)). The key idea is to modify the integrand in equation (1) as \( \tilde{Y}_a(\varepsilon, \theta) = Y_a(\varepsilon, \theta) + \Delta_a(\varepsilon, \theta) \). By construction, we have that \( \mathbb{E}_{\varepsilon \sim p}[\tilde{Y}_a(\varepsilon, \theta)] = \mathbb{E}_{\varepsilon \sim p}[Y_a(\varepsilon, \theta)] \). While the expectation of \( \tilde{Y}_a \) remains unchanged, its covariance matrix will, in general, not be the same.

On the remaining of these notes, we will focus on the diagonal elements of the covariance matrix of \( \tilde{Y}_a \) for simplicity. We can relate the variance of \( \tilde{Y}_a \) to the variance of \( Y_a \) through the covariance rule for a sum of random variables:

\[
\text{Var}[\tilde{Y}_a] = \text{Var}[Y_a] - 2\mathbb{E}[Y_a \Delta_a] + \mathbb{E}[\Delta_a^2].
\]

As can be immediately seen from the covariance rule (2), the control variate \( \Delta_a \) will be useful iff

\[
2\mathbb{E}[Y_a \Delta_a] + \mathbb{E}[\Delta_a^2] < 0.
\]

Here lies the Achilles heel of control variates: On practice, for most interesting cases, we can only approximate a control variate \( \Delta_a \) up to some residual noise. If we assume the residual noise to be uncorrelated to all quantities and have variance \( \kappa^2 \), the condition in equation (3) becomes even harder to satisfy:

\[
2\mathbb{E}[Y_a \Delta_a] + \mathbb{E}[\Delta_a^2] + \kappa^2 < 0.
\]

2 Particular cases

2.1 Baseline Removal

A case of particular interest to Reinforcement Learning, is when \( \nabla_{\theta_a} J(\theta) \) is of the form \( \nabla_{\theta_a} J(\theta) = \mathbb{E}_{\varepsilon \sim p(\varepsilon | \theta)}[C(\varepsilon) \nabla_{\theta_a} \ln p(\varepsilon | \theta)] \) (i.e. \( Y_a(\varepsilon, \theta) = C(\varepsilon) \nabla_{\theta_a} \ln p(\varepsilon | \theta) \)). In the RL literature this is known as policy gradient (the function \( C(\varepsilon) \) is the cost of "taking an action \( \varepsilon \)"") and in the statistics literature it is known as likelihood ratio.

A popular control variate for this gradient is based on the observation that \( \mathbb{E}_{\varepsilon \sim p(\varepsilon | \theta)}[\nabla_{\theta_a} \ln p(\varepsilon | \theta)] = 0 \forall \theta \). Thus, \( \Delta_a(\varepsilon, \theta) = m_a \nabla_{\theta_a} \ln p(\varepsilon | \theta) \) is a valid control variate for any \( m_a \in \mathbb{R} \).

For this to actually be useful we must find a vector \( m_a \) such that \( \text{Var}[\tilde{Y}_a] < \text{Var}[Y_a] \). The condition in equation (3) becomes

\[
\text{Var}[\tilde{Y}_a] = \text{Var}[Y_a] - 2m_a \mathbb{E}[CG_a^2] + m_a^2 \mathbb{E}[G_a^2],
\]

where
where $G_a = \nabla_{\theta_a} \ln p(\varepsilon|\theta)$.

The variance gap above is a convex function of $m_a$ with minimum at

$$m_a^* = -\frac{\mathbb{E}[CG_a^2]}{\mathbb{E}[G_a^2]}.$$ 

The quantity $m_a^*$ is known as optimal baseline in the RL literature. Substituting $m_a^*$ back in (4) we get

$$\text{Var}[\hat{Y}_a] - \text{Var}[Y_a] = -\frac{\mathbb{E}[CG_a^2]^2}{\mathbb{E}[G_a^2]} \leq 0.$$ 

Therefore, the control variate $\Delta_a(\varepsilon, \theta) = -\frac{\mathbb{E}[CG_a^2]}{\mathbb{E}[G_a^2]} \nabla_{\theta_a} \ln p(\varepsilon|\theta)$ is guaranteed to reduce the variance of the original estimator.

However, it is very common in RL to adopt a much simpler control variate: $\Delta_a(\varepsilon, \theta) = \mathbb{E}[C(\varepsilon)] \nabla_{\theta_a} \ln p(\varepsilon|\theta)$. This is known as the average baseline in RL.

Substituting this in equation (4) we get

$$\text{Var}[\hat{Y}_a] - \text{Var}[Y_a] = \mathbb{E}[C]^2 \mathbb{E}[G_a^2] + \mathbb{E}[C]^2 \mathbb{E}[G_a^2].$$ 

(6) Note from the expression above that there is no guarantee that the average baseline will reduce variance.

### 2.2 Symmetric Densities

Assume that the density $p(\varepsilon)$ in equation (1) is symmetric around the origin, i.e. $p(\varepsilon) = p(-\varepsilon)$. An example of a density with this property is a spherical Gaussian distribution.

In this case, any anti-symmetric function $\Delta_a(\varepsilon, \theta) = -\Delta_a(-\varepsilon, \theta)$ will be a valid control variate since $\mathbb{E}_{\varepsilon \sim p}[\Delta_a(\varepsilon, \theta)] = 0$.

In order to find the anti-symmetric function $\Delta_a(\varepsilon, \theta)$ which yields the lowest variance, we need to solve a constrained optimization over the space of anti-symmetric functions. This can be done by introducing a Lagrange multiplier $\lambda_a(\varepsilon)$ which enforces the anti-symmetry. The unconstrained problem to be minimized with respect to both $\lambda$ and $\Delta$ is then

$$2\mathbb{E}[Y_a \Delta_a] + \mathbb{E}[\Delta_a^2] + \int d\varepsilon \lambda_a(\varepsilon)[\Delta_a(\varepsilon, \theta) + \Delta_a(-\varepsilon, \theta)].$$

The minimum of this Lagrangean can be easily found and is given by:

$$\lambda_a^*(\varepsilon) = \frac{Y_a(\varepsilon) + Y_a(-\varepsilon)}{2},$$

$$\Delta_a^*(\varepsilon) = \frac{Y_a(\varepsilon) - Y_a(-\varepsilon)}{2}.$$ 

There is a curious result in here. Replacing $\Delta_a^*(\varepsilon)$ in $\hat{Y}_a(\varepsilon, \theta)$ gives $\hat{Y}_a(\varepsilon, \theta) = Y_a(\varepsilon, \theta) + \Delta_a(\varepsilon, \theta) = \frac{Y_a(\varepsilon) + Y_a(-\varepsilon)}{2}$.

That is, we have just derived antithetic variables from a control variate. Pretty cool.

However, by computing the corresponding variance gap (2) we conclude that antithetic variables are not guaranteed to reduce variance. In order to reduce the variance we need that $\mathbb{E}[Y_a(\varepsilon)Y_a(-\varepsilon)] < \mathbb{E}[Y_a(\varepsilon)^2]$.
2.3 Generic basis function

If we have a set of functions \( \phi_i(\varepsilon), i = 1\ldots M \) such that \( \mathbb{E}[\phi_i] = 0 \), we can introduce the following control variate

\[
\Delta_a(\varepsilon, \theta) = m_a^T \phi(\varepsilon),
\]

where \( m_a \in \mathbb{R}^M \).

The variance gap (2) associated with this control variate can be written as

\[
\text{Var}[\tilde{Y}_a] - \text{Var}[Y_a] = 2 m_a^T \Psi_a + m_a^T \Omega m_a,
\]

where \( \Psi_a = \mathbb{E}[Y_a(\varepsilon) \phi(\varepsilon)] \) and \( \Omega = \mathbb{E}[\phi(\varepsilon) \phi(\varepsilon)^T] \).

The lowest variance is achieved by the optimal control variate

\[
\Delta_a^*(\varepsilon, \theta) = - \Psi_a^T \Omega^{-1} \phi(\varepsilon).
\]

The optimal control variate \( \Delta_a^*(\varepsilon, \theta) \) results in a variance gap \( \text{Var}[\tilde{Y}_a] - \text{Var}[Y_a] = - \Psi_a^T \Omega^{-1} \Psi_a \).

Therefore, it is sufficient that the matrix \( \Omega^{-1} \) be positive-definite in order to have variance reduction.

3 Summary

While control variates are very generic tools, their usefulness requires a careful case-by-case analysis.

A quick summary of the control variates analyzed is provided in the table below.

<table>
<thead>
<tr>
<th>name</th>
<th>assumptions</th>
<th>guaranteed to reduce variance regardless ( Y )?</th>
</tr>
</thead>
<tbody>
<tr>
<td>optimal baseline</td>
<td>need two expectations</td>
<td>yes</td>
</tr>
<tr>
<td>average baseline</td>
<td>need one expectation</td>
<td>no</td>
</tr>
<tr>
<td>antithetic</td>
<td>symmetric density</td>
<td>no</td>
</tr>
<tr>
<td>basis function</td>
<td>( M ) expectations</td>
<td>no (yes if ( \Omega ) is positive-definite)</td>
</tr>
</tbody>
</table>

Table 1. Summary of the control variates analyzed