

## Expected Improvement

For our final search method, we proceed in the following way. First we note the currently best sampled point  $y_{min} = \min(Y = \{y_1, \dots, y_n\})$ . This point will serve as the threshold we wish to overcome. Let  $x \in \Omega$  be some point, we do not know  $f(x)$ , but given the GPR-model  $\mathbf{G}$ , we have access to an estimate of the value  $\mu(x)$  and uncertainty  $\sigma(x)$ . We leverage these by modelling  $f(x)$  with a normally distributed random variable  $A$ , with mean and spread given by  $\mathbf{G}$ . That is,

$$A \sim N(\mu, \sigma) \quad , \quad f_A(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

To measure the improvement at  $x$ , we introduce  $I$ , which we shall define as,

$$I[A(x)] = \max\{y_{min} - A(x), 0\}.$$

Given some realization of  $A$ , the improvement from  $y_{min}$  is at minimum zero, corresponding to  $A$  being greater than  $y_{min}$ . Otherwise, the improvement will be the distance  $A$  lies below  $y_{min}$ . We now wish to calculate the expected value of the improvement, which we call the expected improvement. Which can be done in the usual way using LOTUS,

$$\begin{aligned} \mathbb{E}[I] &= \int_{-\infty}^{\infty} I(z) f_A(z) dz \\ &= \int_{-\infty}^{\infty} \max\{y_{min} - z, 0\} f_A(z) dz \\ &= \int_{-\infty}^{y_{min}} (y_{min} - z) f_A(z) dz + 0. \end{aligned}$$

Substituting here, we are able to recover the standard normal distribution,

$$= \int_{-\infty}^{\frac{y_{min}-\mu}{\sigma}} (y_{min} - r\mu - \sigma) \theta(r) dr.$$

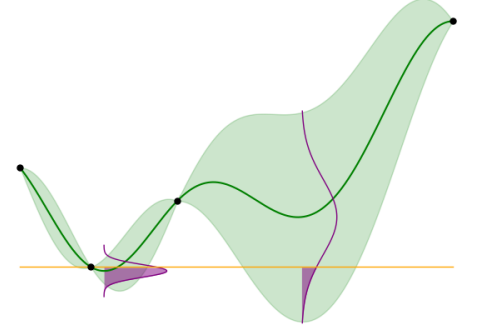
By some fairly tedious integration by parts this can be shown to give,

$$\mathbb{E}[I] = (y_{min} - \mu) \Theta\left(\frac{y_{min} - \mu}{\sigma}\right) + \sigma \theta\left(\frac{y_{min} - \mu}{\sigma}\right).$$

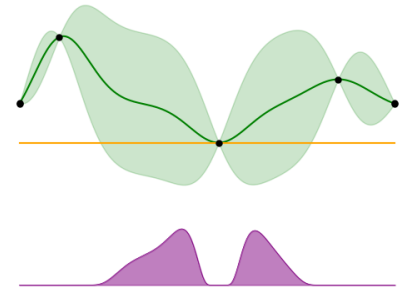
where  $\theta, \Theta$  are the standard normal density and distribution functions. For each  $x \in \Omega$  we can then calculate the expected improvement and select as our next point,

$$x^+ = \arg \max_{x \in \Omega} \mathbb{E}[I(x)].$$

Stated less formally, what we are doing is the following. For each point  $z$  below the threshold  $y_{min}$ , we calculate the improvement  $I(z)$  and weight these with the probability  $f_A(z)$ . Keep in mind that  $f_A(z)$  tends to zero, as  $z$  increases. The expected improvement will therefore be greater when  $f_A$  has more of its mass concentrated below  $y_{min}$ .



**figure:** Expected improvement calculated at two points. Notice, the second gaussian extends further below the EI threshold than the first gaussian. However, the first gaussian has more of its mass concentrated below the threshold. The expected improvement is therefore greater at this point.



**figure:** EI shown for the entire space. Notice, that EI quickly approaches zero, as the surrogate model's uncertainty moves above the threshold.

### 1.0.1 The $\delta$ parameter

Our current algorithm, has a slight problem. In its current form, we are unable to steer the relationship between exploration and exploitation. We can remedy this by adding an additional parameter  $\delta$ . This parameter should be interpreted as moving the threshold we wish to beat,

$$I_\delta[A(x)] = \max\{(y_{min} - \delta) - A(x), 0\}.$$

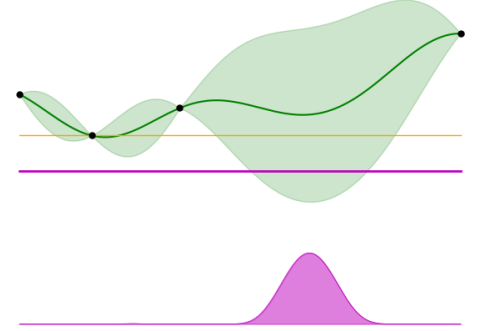
Performing similar calculations as before, we can find the expected improvement to be,

$$\mathbb{E}[I] = ((y_{min} - \delta) - \mu) \Theta\left(\frac{(y_{min} - \delta) - \mu}{\sigma}\right) + \sigma \theta\left(\frac{(y_{min} - \delta) - \mu}{\sigma}\right).$$

The way the parameter  $\delta$  affects the relationship between exploration and exploitation, can be illustrated with the following example. Let us assume we have a situation where,

$$G^{-1}(y_{min}) = \arg \min_{x \in \Omega} \{\mu(x)\}.$$

Corresponding to the situation, where the minimum sampled point is also the minimum of the surrogate model (As in figure 1 in the previous section). In this case, points in the vicinity of  $G^{-1}(y_{min})$ , are almost guaranteed to have positive expected improvement, as points  $x'$  in the vicinity will have values of  $\mu(x')$  close to  $y_{min}$ . This implies that  $f_A(x')$  has a lot its mass concentrated below  $y_{min}$ . Therefore, the model will heavily favor exploitation. The addition of the parameter  $\delta$  (moving the threshold down) penalizes points these points as they, in general, have low uncertainty. In practice, the  $\delta$  is finetuned until a good balance between exploration and exploitation is found.



**figure:** Moving the threshold encourages exploration, as points in the vicinity of  $y_{min}$  are given low uncertainties by the GP. As a result, their probability distributions tend to be tightly concentrated around  $y_{min}$ .