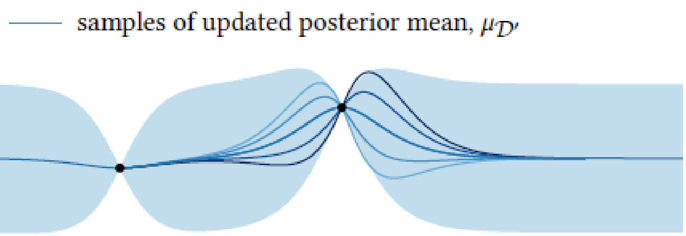
Thanks for the discussion yesterday. I revisited the gradient-related topics for EI and KG, and organized some thoughts regarding the gradient information used during the search for the next point to observe for the outer optimization loop based on the discussion in Chapter 7 of the book (page 129-131)

- Fig. 7.6 showed different samples of the updated posterior mean, derived from sampling from the predictive distribution at the chosen evaluation location by KG (see Fig. 7.5) and conditioning (i.e., when the actual observation has not yet been observed)
    - All of them are plausible with different confidence
    - And we observe that the local optimum can occur on either side of the current best-seen point before being confirmed (to track different locations, we denote this current best-seen point as x_1, and the proposed location by EI/KG as x)
    - *Note*, we know that there must be a local max of the obj. fun. In the neighborhood of the best-seen pt so far, since the posterior mean either go up or down near this point, unless the best-seen pt is actually a local max



- *As noted on p. 130-131,* for KG, making an observation on either side of the current best-seen point can be beneficial, since it will reveal the derivative information around that point
    - Knowing the derivative information around that point is sufficient to have an idea on how the posterior mean function evolves near that region, and no further evaluations are necessary to confirm that belief (of course, it is also related to how much we want to trust the updated posterior mean after observing x)
        - Based on the derivative information, it will always lead to a new max for the updated posterior mean after making the observation, regardless of the location of x (unless the current best-seen x_1 is precisely a local max)
            - And the max of the updated posterior mean (denote as x_1') is **not** necessary at the point we have previously observed (denote as $\boldsymbol{x}$) or at this newly observed location x'

        - By trusting the updated belief, the KG explores more and search globally
            - Instead of sampling around the max of updated posterior mean to confirm the belief, KG utilizes x_1' directly in the formulation when search the next point to observe

- Based on the previous discussions, KG can be better compared to EI, considering the following perspectives
    - For EI, since we only check the observed points $(\boldsymbol{x}, x')$ during inner optimization loop, we ignore the derivative information and the max of the updated posterior mean x_1' → we guess the local max is either on the left or right of the current best that has been observed (either at x_1 or x'), which we hope to be correct; if guessed wrong, we add one more sample to the observed dataset, but the second term in the EI is actually not really updated in terms of the location, which will still be (either at x_1 or x') from the previous iteration (the value will likely to change since the posterior mean function is updated)
        - The credibility of this hope depends on the credibility of the posterior mean, and the predictive distribution, and
            - Similarly, for KG, if the credibility of x_1' is actually low, it may lead to undesired/excessive explorations
        - When x_1' coincides with x', the second term in KG and EI is equivalent in terms of the candidate location (the value will still likely to be different, unless the previously suggested point by KG and EI are identical, which is very unlikely)
- That was what I was referring to during the discussion about the credibility. Look forward to further discussions 😊