GP model

$$y = f^T(x)\beta + M(x), M(x) \sim GP(0, k)$$
$$Y = y + \epsilon, E\epsilon(x) = 0, Var(\epsilon(x)) = r(x)$$

Assume there are a_k replications at covariate x_k,

$$\begin{aligned}\bar{Y}_n &= [\frac{1}{a_1}\sum_{i=1}^{a_1} Y_{1i}, ..., \frac{1}{a_n}\sum_{i=1}^{a_n} Y_{ni}]^T \\ &= [f^T(x_1)\beta + M(x_1) + \bar{\epsilon}_1, ..., f^T(x_n)\beta + M(x_n) + \bar{\epsilon}_n]^T\end{aligned}$$

Let K_n denote the covariance matrix $Cov(M(X_i), M(X_j))_{1\le i,j\le n}$, R_n denote the covariance matrix $diag(\frac{r(X_1)}{a_1}, ..., \frac{r(X_n)}{a_n})$,

$$\Rightarrow Cov(Y(X_i), Y(X_j))_{1\le i,j\le n} = \Sigma_n = K_n + R_n$$

The best linear predictor based on least square:

$$\begin{aligned}E(\lambda_0 + \lambda_1^T\bar{Y}_n - y_0)^2 &= E(\lambda_0 + \lambda_1^T\bar{Y}_n - \lambda_1^T E\bar{Y}_n + \lambda_1^T E\bar{Y}_n - f^T(x_0)\beta - M(x_0))^2 \\ &= E(\lambda_0 + \lambda_1^T E\bar{Y}_n - f^T(x_0)\beta)^2 + E(\lambda_1^T\bar{Y}_n - \lambda_1^T E\bar{Y}_n - M(x_0))^2 \\ &= (\lambda_0 + \lambda_1^T E\bar{Y}_n - f^T(x_0)\beta)^2 + \lambda_1^T Var(\bar{Y}_n)\lambda_1 + k(x_0, x_0) - 2\lambda_1^T Cov(\bar{Y}_n, M(x_0)) \\ &= (\lambda_0 + \lambda_1^T E\bar{Y}_n - f^T(x_0)\beta)^2 + \lambda_1^T\Sigma_n\lambda_1 + k(x_0, x_0) - 2\lambda_1^T k_n(x_0)\end{aligned}$$

$$\begin{aligned}minE(\lambda_0 + \lambda_1^T\bar{Y}_n - y_0)^2 &\Rightarrow \lambda_1 = \Sigma_n^{-1}k_n(x_0), \lambda_0 = f^T(x_0)\beta - k_n^T(x_0)\Sigma_n^{-1}E\bar{Y}_n \\ \hat{y}_0 &= \lambda_0 + \lambda_1^T\bar{Y}_n \\ &= f^T(x_0)\beta + k_n^T(x_0)\Sigma_n^{-1}(\bar{Y}_n - F_n\beta)\end{aligned}$$

The MSE of this estimator:

$$MSE(x_0) = E(\lambda_0 + \lambda_1^T\bar{Y}_n - y_0)^2 = k(x_0, x_0) - k_n^T(x_0)\Sigma_n^{-1}k_n(x_0)$$

Recursive relationship between MSE
Let $\sigma_n^2(x) = k(x,x) + r(x) - k_n(x)^T\Sigma_n^{-1}k_n(x)$, $k_n(x) = [k(x, X_1), ..., k(x, X_n)]^T$,

Proposition 1.
In the nth step of a sequential design, if X_{n+1} is not a replication, X_{n+1} should be the maximum point of $\Delta_n = \int_D \frac{1}{\sigma_n^2(X_{n+1})}(k(x, X_{n+1}) - k_n^T(x)\Sigma_n^{-1}k_n(X_{n+1}))^2 dP(x)$.
Proof:
Let $I_n = \int_D MSE_n(x|X_1, ..., X_n)dP(x)$ denote the integrated MSE, by lemma.1,
$I_{n+1} = I_n - \int_D \frac{1}{\sigma_n^2(X_{n+1})}(k(x, X_{n+1}) - k_n^T(x)\Sigma_n^{-1}k_n(X_{n+1}))^2 dP(x)$

Lemma 1.
If X_{n+1} is not a replication, $MSE_{n+1}(x|X_1, ..., X_n, X_{n+1}) = MSE_n(x|X_1, ..., X_n) -$

$\frac{1}{\sigma_n^2(X_{n+1})}(k(x, X_{n+1}) - k_n^T(x)\Sigma_n^{-1}k_n(X_{n+1}))^2$
Proof:

$$\Sigma_n = \begin{bmatrix} k(X_1, X_1) + \frac{r(X_1)}{m_1} & ... & k(X_1, X_n) \\ ... & ... & ... \\ k(X_n, X_1) & ... & k(X_n, X_n) + \frac{r(X_n)}{m_n} \end{bmatrix}$$

$$\Sigma_{n+1} = \begin{bmatrix} \Sigma_n & k_n(X_{n+1}) \\ k_n^T(X_{n+1}) & k(X_{n+1}, X_{n+1}) + r(X_{n+1}) \end{bmatrix}$$

According to Block matrix inverse formula and WoodBury Matrix Identity,

$$\begin{aligned}\Sigma_{n+1}^{-1} &= \begin{bmatrix} (\Sigma_n - k_n(X_{n+1})(k(X_{n+1}, X_{n+1}) + r(X_{n+1}))^{-1}k_n^T(X_{n+1}))^{-1} & -\frac{1}{\sigma_n^2(X_{n+1})}\Sigma_n^{-1}k_n(X_{n+1}) \\ -\frac{1}{\sigma_n^2(X_{n+1})}k_n^T(X_{n+1})\Sigma_n^{-1} & \frac{1}{\sigma_n^2(X_{n+1})} \end{bmatrix} \\ &= \begin{bmatrix} \Sigma_n^{-1} + \frac{1}{\sigma_n^2(X_{n+1})}\Sigma_n^{-1}k_n(X_{n+1})k_n^T(X_{n+1})\Sigma_n^{-1} & -\frac{1}{\sigma_n^2(X_{n+1})}\Sigma_n^{-1}k_n(X_{n+1}) \\ -\frac{1}{\sigma_n^2(X_{n+1})}k_n^T(X_{n+1})\Sigma_n^{-1} & \frac{1}{\sigma_n^2(X_{n+1})} \end{bmatrix}\end{aligned}$$

Therefore,

$$\begin{aligned} MSE_{n+1}(x|X_1, ..., X_{n+1}) &= k(x, x) - k_{n+1}^T(x)\Sigma_{n+1}^{-1}k_{n+1}(x) \\ &= k(x, x) - [k_n^T(x) \qquad k(x, X_{n+1})]\Sigma_{n+1}^{-1}[k_n^T(x) \qquad k(x, X_{n+1})]^T \\ &= k(x, x) - k_n^T(x)(\Sigma_n^{-1} + \frac{1}{\sigma_n^2(X_{n+1})}\Sigma_n^{-1}k_n(X_{n+1})k_n^T(X_{n+1})\Sigma_n^{-1})k_n(x) \\ &- 2k_n^T(x)\frac{1}{\sigma_n^2(X_{n+1})}\Sigma_n^{-1}k_n(X_{n+1})k(x, X_{n+1}) + \frac{k^2(x, X_{n+1})}{\sigma_n^2(X_{n+1})} \\ &= k(x, x) - k_n^T(x)\Sigma_n^{-1}k_n(x) - \frac{1}{\sigma_n^2(X_{n+1})}(k(x, X_{n+1}) - k_n^T(x)\Sigma_n^{-1}k_n(X_{n+1}))^2 \\ &= MSE_n(x|X_1, ..., X_n) - \frac{1}{\sigma_n^2(X_{n+1})}(k(x, X_{n+1}) - k_n^T(x)\Sigma_n^{-1}k_n(X_{n+1}))^2 \end{aligned}$$

Lemma 2.
If X_{n+1} is a replication, suppose $X_{N+1} = x_k, 1 \leq k \leq n$, then

$$\Delta_n = k_n^T(x)\frac{(K_n^{-1})_{*,k}(K_n^{-1})_{k,*}}{a_k(a_k + 1)/r(x_k) - (K_n^{-1})_{k,k}}k_n(x)$$

Proposition 2.
Let d denote the dimension of X, $h(x)$ denote the density of X, $g(x) = k(x, X_{n+1}) - k_n^T(x)\Sigma_n^{-1}k_n(X_{n+1})$, $H = Hess_{log|g(x)|}(X_{n+1})$.

$$\Delta_n = \frac{1}{\sigma_n^2(X_{n+1})}\int_D g^2(x)h(x)dx \approx \frac{\pi^{d/2}}{\sigma_n^2(X_{n+1})}h(X_{n+1})g^2(X_{n+1})|H|^{-1/2}$$

Proof:

By Laplace method,

$$\begin{aligned}
\int_D g^2(x)h(x)dx &= \int_D h(x)e^{nlog^{g^{2/n}(x)}}dx \\
&\approx (\frac{2\pi}{n})^{\frac{d}{2}} h(x_{max})g^2(x_{max})|Hess_{\frac{1}{n}log^{g^2(x)}}(x_{max})|^{-1/2} \\
&= \pi^{d/2}h(x_{max})g^2(x_{max})|Hess_{log^{|g(x)|}}(x_{max})|^{-1/2}
\end{aligned}$$

x_{max} is the maximum point of $log^{g^{2/n}(x)}$, i.e. x_{max} is the maximum point of $|g(x)|$. As for the x_{max}, by Cauchy–Schwarz inequality,

$$\begin{aligned}
|g(x)| &= |< k(x,*), k(X_{n+1},*) - k_n^T(X_{n+1})\Sigma_n^{-1}[k(X_1,*),...,k(X_n,*)]^T >_{H_k}| \\
&\le \sqrt{k(x,x)(k(X_{n+1},X_{n+1}) - k_n^T(X_{n+1})\Sigma_n^{-1}k_n(X_{n+1}))}
\end{aligned}$$

When equality holds, $x = X_{n+1}$.

Example 1. Gaussian Kernel

$k(x_1,x_2) = \nu e^{-\frac{||x_1-x_2||^2}{\theta}}, x_i = [x_{i1},...,x_{id}]^T$

In this case, for $1 \le i \le d$,

$$\begin{aligned}
g(x) &= k(x,X_{n+1}) - k_n^T(x)\Sigma_n^{-1}k_n(X_{n+1}) \\
\frac{\partial g(x)}{\partial x_i} &= -\frac{2}{\theta}(x_i - X_{n+1,i})k(x,X_{n+1}) + \frac{2}{\theta}k_n^T(X_{n+1})\Sigma_n^{-1}\begin{bmatrix} x_i - X_{1i} & & \\ & \ddots & \\ & & x_i - X_{ni}\end{bmatrix} k_n(x) \\
&= -\frac{2}{\theta}(x_i - X_{n+1,i})k(x,X_{n+1}) + \frac{2}{\theta}k_n^T(X_{n+1})\Sigma_n^{-1}diag(x_i - X_{ti})_{1\le t\le n}k_n(x) \\
\frac{\partial^2 g(x)}{\partial^2 x_i} &= -\frac{2}{\theta}k(x,X_{n+1}) + \frac{4}{\theta^2}(x_i - X_{n+1,i})^2k(x,X_{n+1}) + \frac{2}{\theta}k_n^T(x)\Sigma_n^{-1}k_n(X_{n+1}) \\
&- \frac{4}{\theta^2}k_n^T(X_{n+1})\Sigma_n^{-1}diag(x_i - X_{ti})^2_{1\le t\le n}k_n(x) \\
\frac{\partial^2 g(x)}{\partial x_i \partial x_j} &= \frac{4}{\theta^2}(x_i - X_{n+1,i})(x_j - X_{n+1,j})k(x,X_{n+1}) \\
&- \frac{4}{\theta^2}k_n^T(X_{n+1})\Sigma_n^{-1}diag((x_i - X_{ti})(x_j - X_{tj}))_{1\le t\le n}k_n(x)
\end{aligned}$$

$$Hess_{log^{|g(x)|}}(x) = (h_{ij})_{1\le i,j\le d}$$

$$h_{ij} = \begin{cases} \frac{1}{g(x)}\frac{\partial^2 g(x)}{\partial x_i \partial x_j} - \frac{1}{g^2(x)}\frac{\partial g(x)}{\partial x_i}\frac{\partial g(x)}{\partial x_j}, i \ne j \\ \frac{1}{g(x)}\frac{\partial^2 g(x)}{\partial^2 x_i} - \frac{1}{g^2(x)}(\frac{\partial g(x)}{\partial x_i})^2, i = j \end{cases}$$

Algorithm 1 Sequential design for homoskedastic GP

Input: $X_n, Y_n, h(x), step$

Fit initial parameters θ, ν, σ^2 by MLE

$i \leftarrow 0$

while $i < step$ **do**

  $x_{n+1} \leftarrow argmax \Delta_n$

  $y_{n+1} \leftarrow Y(x_{n+1})$

  $X_n \leftarrow rbind(X_n, x_{n+1})$

  $Y_n \leftarrow rbind(Y_n, y_{n+1})$

  update θ, ν, σ^2

end while

return X_n, Y_n