

Student Info

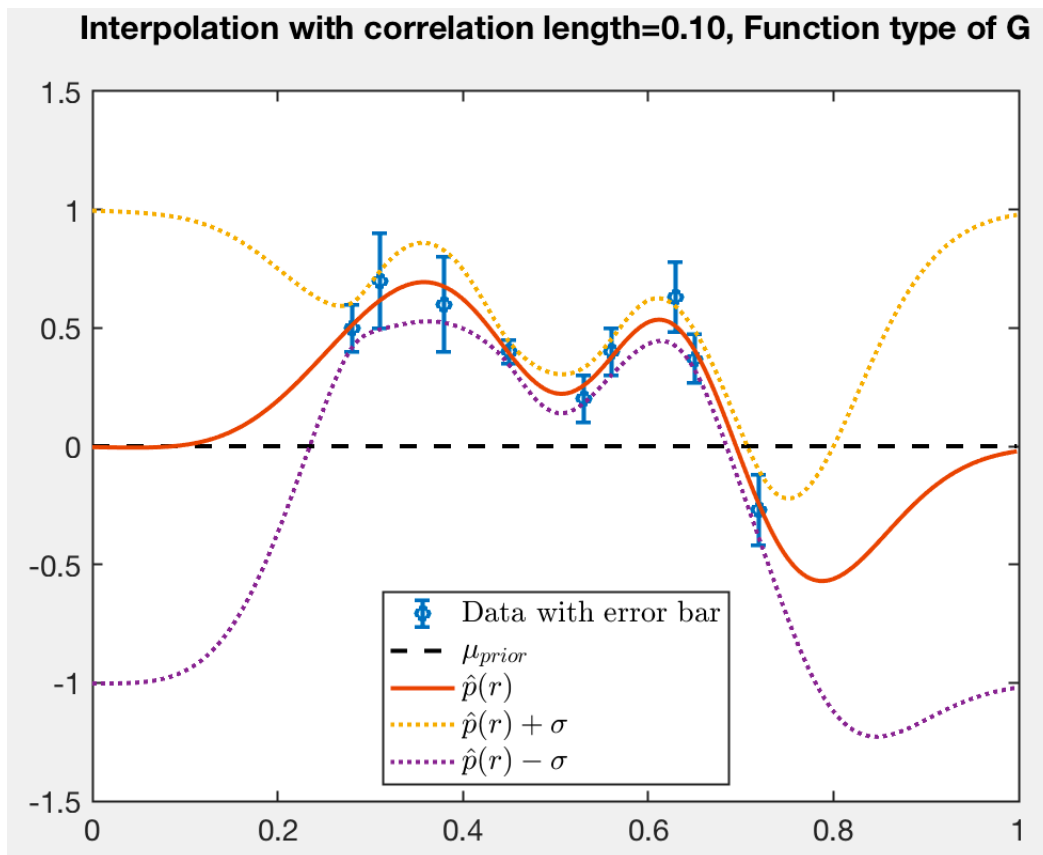
Jingyi Zhuang

jz2907@columbia.edu

Homework 10

A.

The plot:



Script:

```
% initiate variables
```

```

clearvars;
load('QMDA_HW_10.mat');

% Define the matrix
M = 500;
N = length(d);
m = [linspace(0,1,M+1)]';
m = m(1:end-1);
mean_prior = 0;
s_prior = 1;
l_corr = 0.1;
type_corr = 'G';

% sdeve = zeros(N,1);

sig_e = diag(sdeve.^2);

G = zeros(N,M);

for i = 1:1:N
    [r,c] = find(m==xd(i));
    G(i,r) = 1;
end

sig_prior = zeros(M,M);

for i = 1:1:M
    for j = 1:1:M
        sig_prior(i,j) = acvf(m(i)-
m(j),s_prior,l_corr,type_corr);
    end
end

```

```

end

u_prior = mean_prior * ones(M,1);
d_prior = G * u_prior;
A = sig_prior * G';
B = G * A;

u_post = u_prior + A / (sig_e + B) * (d - d_prior);
sig_post = sig_prior - A / (sig_e + B) * A';
sig_post_sd = sqrt(diag(sig_post));

figure(1);
clf;

% Plot the raw data
errorbar(xd,d,sdve,'LineWidth',2,'LineStyle','none','Marker',
,'o');
set(gca,'FontSize',14,'LineWidth',1);
hold on;

% Plot the u_prior
plot(m,u_prior,"k--","LineWidth",2);

% Plot the interpolation result

plot(m,u_post,"LineWidth",2);
plot(m,u_post + sig_post_sd ,"LineWidth",2,"LineStyle",":");
plot(m,u_post - sig_post_sd ,"LineWidth",2,"LineStyle",":");

% Some plot settings
ylim([-1.5 1.5]);
title(sprintf('Interpolation with correlation length=%.2f,
Function type of %s\n',l_corr,type_corr));

```

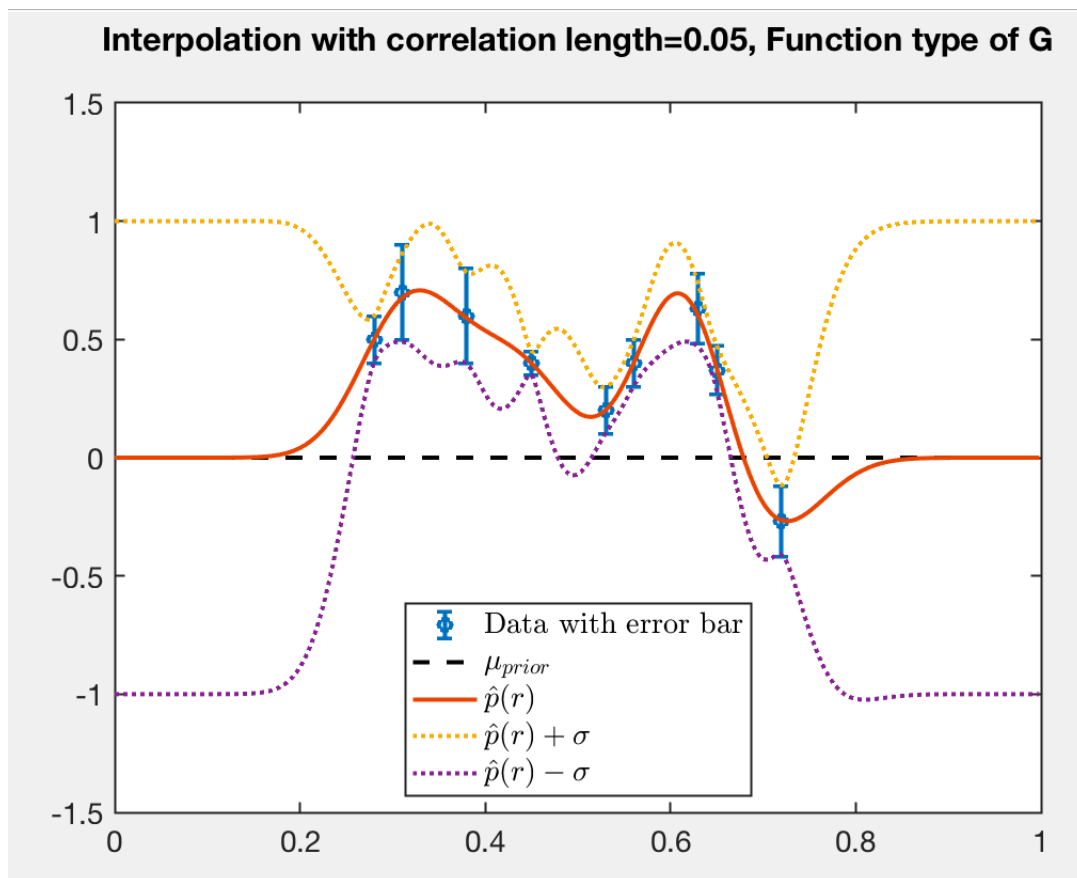
```

legs = legend('Data with error bar', '$\mu_{prior}$',
'$\hat{p}(r)$', '$\hat{p}(r) + \sigma$', '$\hat{p}(r) - \sigma$');
set(legs, 'Interpreter', 'latex');
set(legs, 'FontSize', 14, 'Location', 'south');

```

B.

Change the correlation length to 0.05:



The changes / parameter settings with respect to the script in (A) is:

```

l_corr = 0.05;
type_corr = 'G';

```

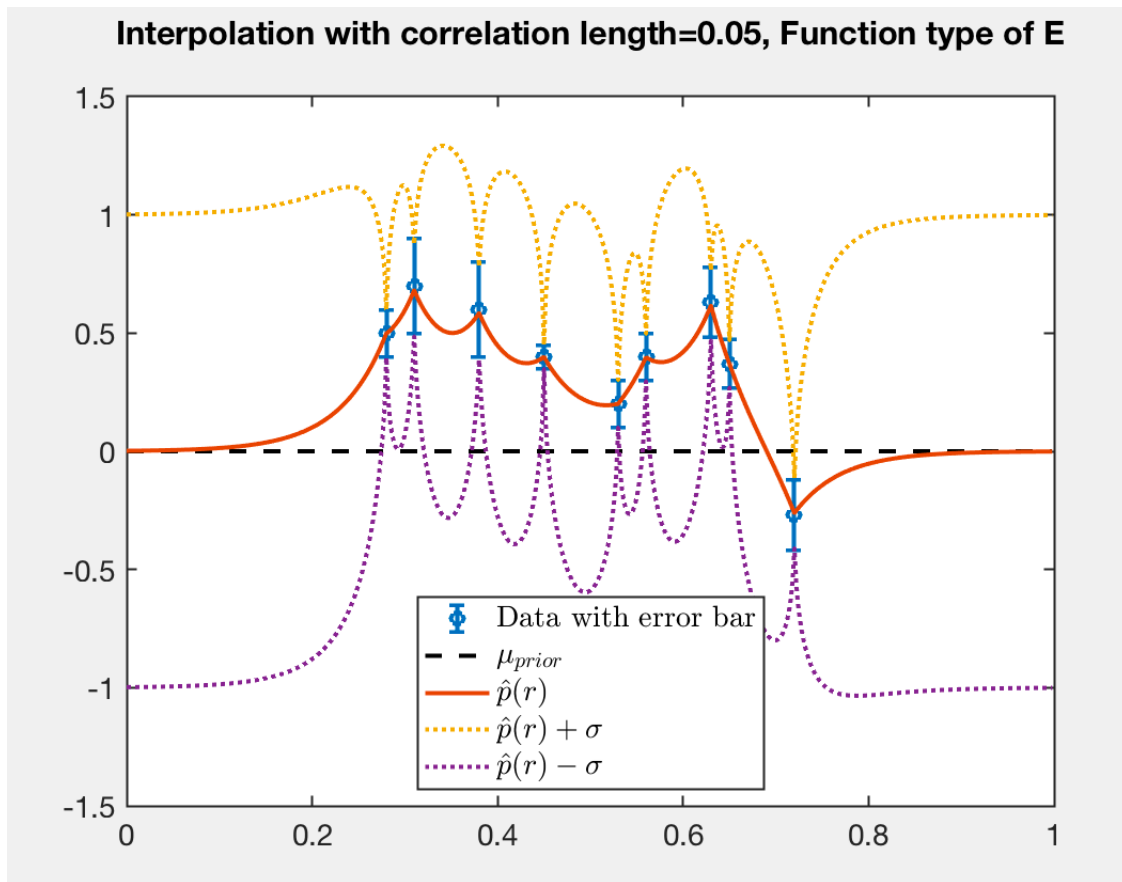
By changing the correlation length to 0.05, the interpolated curve is bended more rapidly, e.g. from $x = 0$ to $x = 0.3$, $l_{corr} = 0.1$ will make the bending started at $x = 0.1$, while $l_{corr} = 0.05$ makes the bending started at $x = 0.2$. The posterior mean is constrained by the prior mean and/or the data point. So, at the data point, the posterior mean is mainly controlled by the data value; while off the data point, the smaller correlation length will lead to more control by the prior mean.

Also similarly for the posterior variance, it's constrained by the prior variance and/or the measurement errors. When it's at the data point, the measurement errors dominate the posterior variance. When it's off the data point, the prior variance will get into the posterior variance. In the case of shorter correlation length, the measurement errors affect smaller range around the data point across the x-axis, so the posterior variance will tend to be more of the prior variance when it's more off the data point.

When $l_{corr} = 0.1$, the interpolated red line $\hat{p}(r)$ goes through the data uncertainty region, while the $\hat{p}(r) \pm \sigma$ do not always go cross the data error bar regions. However when $l_{corr} = 0.05$, the allowed variability is smaller, the interpolated data at $x_i^m = x_j^d$ are mostly going through the observed data points, and the $\hat{p}(r) \pm \sigma$ of this setting also go almost exactly on the error bars. That's because the longer correlation length will make the measurement errors of neighbor data points affect each other more than the shorter correlation length.

C.

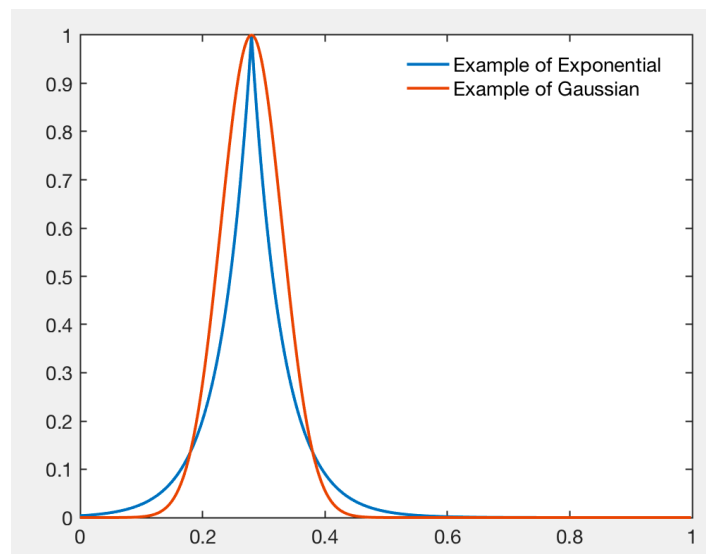
Change the type from Gaussian to Exponential:



The changes / parameter settings with respect to the script in (A) is:

```
l_corr = 0.05;
type_corr = 'E';
```

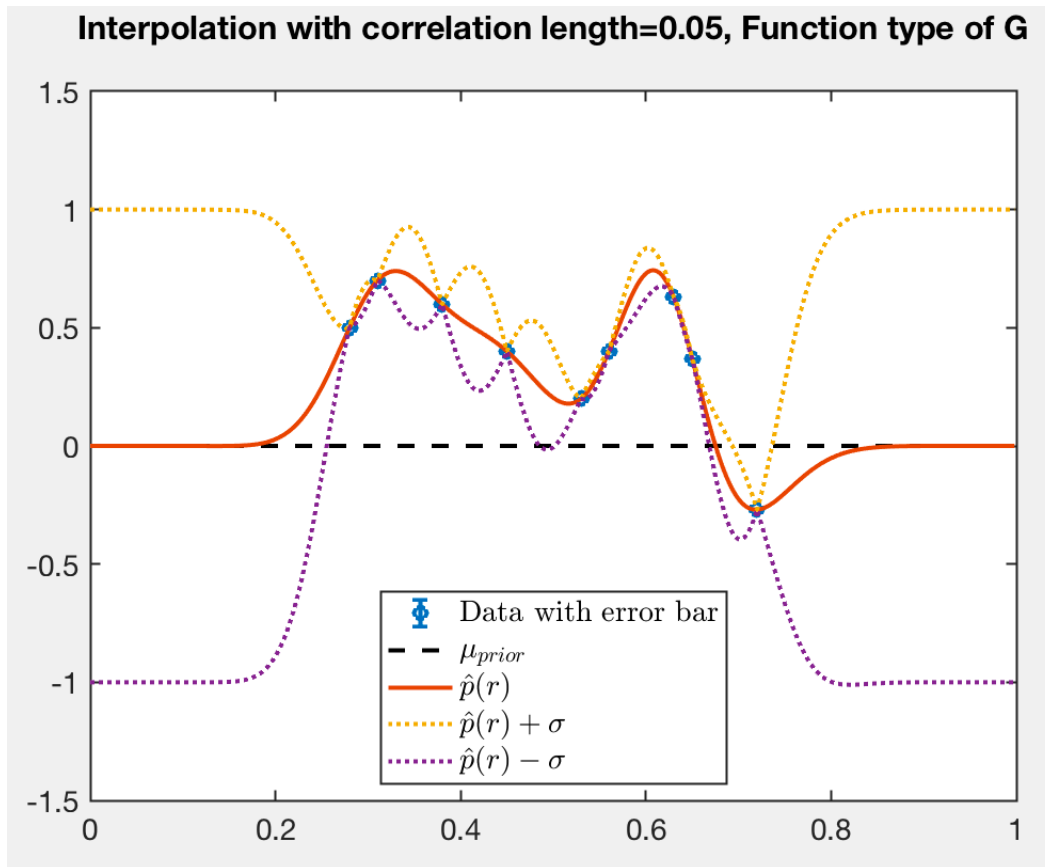
To better describe this figure, I will need an illustration to show the shape of exponential and gaussian function:



It's shown that, when leaving the data point, exponential function shows a much more rapid decrease, while the gaussian is kind of smoothed around the data point. Thus, when leaving the data point, the posterior mean is more urgent to get close to the prior mean in the exponential function than in the gaussian function. Same for the posterior variance, it is more urgent to approach the prior variance, with sharp vertices, in the exponential function. The gaussian function will give smoother interpolation.

D.

No error bar:



The changes / parameter settings with respect to the script in (A) is:

```
l_corr = 0.05;  
type_corr = 'G';  
  
sdeve = zeros(N,1);  
sig_e = diag(sdeve.^2);
```

The interpolated posterior mean $\hat{p}(r)$ is going through each data point, as well as the $\hat{p}(r) \pm \sigma$, when there are zero measurement errors.