P. Koumoutsakos, P. Hadjidoukas
ETH Zentrum, HG D1.1/G3
CH-8092 Zürich

Spring semester 2018

Set 01 - Gaussian, OpenMP and MPI reminders

Issued: February 26, 2018

Question 1: Probability Theory Reminders

A variable with gaussian distribution $X \sim \mathcal{N}(\mu, \sigma)$ has a probability distribution function given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

a) Show that the mean of X is given by $\mathbb{E}[X] = \mu$.

b) Show that the variance of X is given by $\mathbb{E}[(X-\mu)^2] = \sigma^2$.

c) Let $X \sim \mathcal{N}(\mu_X, \sigma_X)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y)$. Show that $Z = X + Y \sim \mathcal{N}(\mu_Z, \sigma_Z)$, with $\mu_Z = \mu_x + \mu_Y$ and $\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2$.
Hint: The pdf of the sum of two independent variables X and Y with probability distribution functions f_X and f_Y is given by the convolution

$$f_{X+Y}(y) = \int_{-\infty}^{+\infty} f_Y(y-x) f_X(x) dx.$$

Question 2: Statistics of Diffusion Solver

In `coding_2_diffusion_hybrid/diffusion2d_mpi.cpp` you find an MPI-based solver for the 2D diffusion problem.

The provided code includes a sequential function `compute_histogram_seq()` which computes and prints the histogram of the density values of the local subdomain using M (=10) bins.

a) Provide a hybrid MPI + OpenMP parallel implementation of the above function in `compute_histogram_hybrid()`. Process with rank 0 must print the combined histogram for all density values.

b) Extend the rest of the 2D diffusion code so as to provide a complete hybrid MPI + OpenMP parallel implementation.

Write your solution in `coding_2_diffusion_hybrid/diffusion2d_hybrid.cpp` and adapt the Makefile in the same folder to compile your program using make.

Question 3: Monte Carlo and OpenMP recap

a) Posterior calculation

In Bayesian uncertainty quantification, we are interested in *posterior distributions* of model parameters:

$$f(\beta \mid \mathcal{D}) = \frac{f(\mathcal{D} \mid \beta) f(\beta)}{\int_{\Omega_\beta} f(\mathcal{D} \mid \beta) f(\beta) \mathrm{d}\beta}, \tag{1}$$

where $\mathcal{D}$ denotes a set of observed data, and β model parameters, To use the Bayesian formula, the denominator integral needs to be evaluated, which is in most cases analytically intractable.

Your task is to study and implement the following integration methods, and evaluate the approximation to following integral:

$$\int_a^b f(x) \mathrm{d}x. \tag{2}$$

Test the methods on these functions:

- $f(x) = -(x-1)(x+1)^2, x \in [-1, 1]$
- $f(x) = -\exp(-x), x \in [0, \infty]$

i) Trapezoidal integration

$$\int_a^b f(x) x \approx h \left(\frac{1}{2} f(x_0) + f(x_1) + \ldots + f(x_{N-1}) + \frac{1}{2} f(x_N) \right), \tag{3}$$

$$h = \frac{b-a}{N}, x_i = a + ih, i = 0 \ldots N, \tag{4}$$

where N is the number of intervals we divide the (b, a) into.

Tasks:

1) Implement the serial version of the method, and verify your code by comparing the results with those of the analytical solutions.
2) Parallelize using OpenMP and report the speedup.

ii) Monte Carlo integration

Instead of implementing a grid based approach, we use the following idea: suppose $X \sim Unif(a, b)$. Then:

$$\mu_f = \mathbb{E}[f] = \frac{1}{b-a} \int_a^b f(x) \mathrm{d}x \to \int_a^b f(x) \mathrm{d}x = (b-a)\mu_f \tag{5}$$

An unbiased estimator for the mean is:

$$\hat{\mu}_f = \frac{1}{N} \sum_{i=1}^{n} f(x_i) \to \int_a^b f(x) \mathrm{d}x \approx (b-a)\hat{\mu}_f, \tag{6}$$

where $\{x_i\}$ are independent realizations of the random variable X.

1) Is the estimator of the integral unbiased?

2) What is the variance of the integral estimator?
3) Implement a serial version of the estimator and verify your code by comparing the results with those of the analytical solutions. Compare error convergence with the trapezoidal method for different values of N.
4) Parallelize using OpenMP. Do you need to use barriers? Why or why not? Report the speedup.
5) Bonus: can you exploit the knowledge that our integral is of the form $\int a(x)b(x)\mathrm{d}x$?