IGL PROJECT FINAL REPORT - OPTIMAL PARTITIONS

J. ROSENBLATT, IGL FACULTY ADVISOR

ABSTRACT. The Riemann integral can be approximated using partitions and a rule for assigning weighted sums of the function at points determined by the partition. Approximation methods commonly used include endpoint rules, the midpoint rule, the trapezoid rule, Simpson’s rule, and other quadrature methods. The rate of approximation depends to a large degree on the rule being used and the smoothness of the function, but it also depends on the partition. We consider some cases of optimal partitions and develop methods of computing the partition and getting information about the distribution of the points in the optimal partition.

1. INTRODUCTION

There is an extensive literature on optimal approximation, extending from classical approximation theory to modern image processing. So we give only those that we know are directly related to the questions in this report. Despite all of this literature, it has been difficult to find articles about optimal approximation, and the distribution of the parameters that provide the optimal approximation, that use readily understandable, rigorous arguments.

First consider a very basic case of optimal approximation of the Riemann integral. Take a bounded Riemann integrable function $f$ on $[0, 1]$ and choose a partition $P_n = \{x_1, \ldots, x_n\}$ of $[0, 1]$ containing $n$ distinct points. We assume here that $0 = x_0 < x_1 < \cdots < x_n < x_{n+1} = 1$. Use the left endpoint rule $E(f, 0, 1, P_n) = \sum_{k=0}^{n} f(x_k)(x_{k+1} - x_k)$ to obtain an approximation of $\int_0^1 f(t) \, dt$. We want to know how to choose the partition $P_n^\#$ to achieve (close to) the smallest possible error $\int_0^1 f(t) \, dt - E(f, 0, 1, P_n)$ over all $n$ point partitions $P_n$. We would like to answer the following questions, in cases like this one and for more general methods of approximating the integral:

1. When does the optimal partition exist and when is it unique? How do we determine the optimal partition?
2. What is the overall error for the numerical approximation of the integral using the optimal partition as $n$ tends to $\infty$?
3. How are the points of $P_n^\#$ distributed as $n$ goes to $\infty$?

Even if $f$ is strictly increasing and continuous, the answers to these questions are not all obvious. We can answer these question at least when the function is also continuously differentiable.

To understand some of the issues here, let $f$ be continuous and consider the lower Riemann sums $L(f, 0, 1, P_n) = \sum_{k=0}^{n} \left( \min_{x \in [x_k, x_{k+1}]} f(x) \right) (x_{k+1} - x_k)$. If $f$ were non-decreasing, this would of course be the left endpoint rule above. Assuming that the mesh of $P_n$, that

\textit{Date: May, 2012.}
is the maximum value of $x_{k+1} - x_k$, $k = 0, \ldots, n$, tends to zero, we have $L(f, 0, 1, P_n) \to \int_0^1 f(t) \, dt$ as $n \to \infty$. If the mesh tends to zero slowly, then this convergence will be slow too. But if $f$ is continuously differentiable, and one chooses the uniform partition $P_n^U$, where the values $x_{k+1} - x_k = \frac{1}{n+1}$ for all $k$, then the error $|\int_0^1 f(t) \, dt - L(f, 0, 1, P_n^U)|$ is bounded by $\max_{x \in [0,1]} |f'(x)|/n$. One might think that one could get an even better rate of approximation by choosing the partition more specifically with the function in mind. However, Tasaki [2] computed the actual error, using the optimal partition $P_n^#$ for this rule, and showed that for continuously differentiable functions,

$$\lim_{n \to \infty} n \left| \int_0^1 f(t) \, dt - L(f, 0, 1, P_n^#) \right| = \frac{1}{2} \left( \int_0^1 |f'(t)|^{1/2} \, dt \right)^2.$$ 

In addition, Tasaki [2] obtains a result like this for the trapezoid rule. Actually, both of these results follow from the general method in McClure [1].

We would like to determine explicitly the optimal partition $P_n^#$. But while we can write down recursive formulas and do some calculations in this direction, it seems to be difficult to determine this partition precisely in general. However, we can possibly determine its distribution asymptotically as $n$ tends to $\infty$. It is intuitively clear that in regions where $f$ is changing quickly, one has to put more of the points of $P_n^#$ than in other regions, and so the distribution of $P_n^#$ must relate in some fashion to the derivative(s) of $f$. We propose that the best way to study the distribution of a choice of $P_n$ is to consider the probability measure $\nu(P_n) = \nu_n = \frac{1}{n} \sum_{k=1}^n \delta_{x_k}$ where $\delta_x$ denotes the point mass measure at $x$. The question is: does $(\nu(P_n^#))$ converge weakly? That is, does the limit $\Lambda(h) = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^n h(x_k^#)$ exist for all continuous functions $h : [0,1] \to \mathbb{R}$? If this weak limit does exist, then $\Lambda(h) = \int_{[0,1]} h \, d\nu$ for some probability measure $\nu$ on $[0,1]$. So the follow up question is: how is $\nu$ related to $f$? We can use McClure [1] directly to answer this question in some cases, for instance the cases considered by Tasaki [2]. For example, using the lower Riemann integral estimates $L(f, 0, 1, P_n^#)$, when $f$ is continuously differentiable, the weak limit $\nu$ of $\nu(P_n^#)$ is given by

$$\int_{[0,1]} h(t) \, d\nu(t) = \frac{1}{I} \left( \int_0^1 h(t)|f'(t)|^{1/2} \, dt \right),$$

where $I$ is the normalizing factor $I = \int_0^1 |f'(t)|^{1/2} \, dt$.

2. Transforming optimal partitions to uniform partitions

We have indicated that it is difficult to explicitly compute the optimal partitions even in fairly simple cases. But we know that we can compute the asymptotic distribution of these points in some cases. We can also see from this what transform needs to be applied to convert the optimal partitions into ones that are asymptotically uniformly distributed. We illustrate these issues here with the left-hand rule in general, and for some specific functions.

First, suppose $f$ is differentiable on $\mathbb{R}$. Assume that $f' > 0$ and so $f$ is strictly increasing. Take $P_n^#$ to be a partition of $[0,1]$ which maximizes the left endpoint Riemann
sum $L(\mathcal{P}_n) = \sum_{k=0}^{n} f(x_k)(x_{k+1} - x_k)$ i.e. minimizes the error in approximating the Riemann integral by the lower Riemann sum. The optimal partition must have the total differential $DL(f) = 0$ i.e. $\frac{\partial L}{\partial x_i} = 0$ for all $i = 1, \ldots, n$. This means that we have the following equations, for $i = 1, \ldots, n$:

$$f(x_{i-1}) + f'(x_i)(x_{i+1} - x_i) - f(x_i) = 0.$$  

This formula does not immediately determine $x_1$; that value is only determined in the end using the fact that $x_{n+1} = 1$.

**Remark 2.1.** In particular, say $f$ in linear, given by $f(x) = bx$ where $b > 0$. Then

$$bx_{i-1} + b(x_{i+1} - x_i) - bx_i = 0$$

for all $i = 1, \ldots, n$. So $x_{i+1} - x_i = x_i - x_{i-1}$ for all $i = 1, \ldots, n$. This means that the points $x_i$ are uniformly distributed in $[0, 1]$. So in this case the asymptotic distribution of $\mathcal{P}_n$ is the usual Lebesgue measure on $[0, 1]$. This is certainly what we expected from our earlier computations.

**Remark 2.2.** Moreover, in some cases, instead of having to optimize the choice of the partition points in aggregate, which is generally a computationally slow process, one can use the recursive formula to get a fairly efficient algorithm. Consider the case of powers $f(x) = x^d$. Take an optimal partition $\mathcal{P}^\#_n = \{x_1(n), \ldots, x_n(n)\}$ with $n$ points. Then our recurrence formula determining $\mathcal{P}^\#_n$ is

$$x_{i+1}(n) - x_i(n) = \frac{x_i^d(n) - x_{i-1}^d(n)}{dx_i^d-1(n)}.$$  

For $i, 2 \leq i \leq n$, in terms of the ratios $r_i(n)$ defined by $r_i(n) = x_{i+1}(n)/x_i(n)$, this formula becomes

$$r_i(n) - 1 = \frac{1}{dx_i(n)} \left( x_i(n) - \frac{x_{i-1}^d(n)}{r_{i-1}^d(n)} \right) = \frac{1}{d} \left( 1 - \frac{1}{r_{i-1}^d(n)} \right).$$

That is,

$$r_i(n) = \frac{d + 1}{d} - \frac{1}{dr_{i-1}^d(n)}.$$  

We can extend this formula to $i, 1 \leq i \leq n$ by letting $r_1(n) = (d + 1)/d$. This is consistent with the recurrence formula for $r_i(n)$ because we can formally take $r_0(n) = x_1(n)/x_0(n) = x_1(n)/0 = \infty$ and $(1/dr_0^d(n)) = 0$. Now notice that we have for $i = 1, \ldots, n$,

$$x_i(n) = 1/\prod_{j=i}^{n} r_j(n).$$

The advantage of using this formula to compute the points $x_i(n)$ in the optimal partition $\mathcal{P}^\#_n$ is that when computing the points in $\mathcal{P}^\#_{n+1}$, the formulas give $r_i(n + 1) = r_i(n)$ for $i = 1, \ldots, n$ and one need only compute the additional value

$$r_{n+1}(n + 1) = \frac{d + 1}{d} - \frac{1}{dr_{n+1}(n)^d}.$$  

Also, the new partition points $x_i(n + 1)$ are $x_i(n)/r_{n+1}(n + 1)$ for $i = 1, \ldots, n$, and the one additional point $x_{n+1}(n + 1)$ is $1/r_{n+1}(n + 1)$. These recursion formulas can be easily implemented to give quickly all the partition points for large values of $n$. 
Because we know the asymptotic distribution of $\nu(P^*_n)$, we can actually see what transform to use that would convert $P^*_n$ into an asymptotically uniform partition. An interesting aspect of this transform is that it does not convert $P^*_n$ to a precisely uniform partition.

The transform that we need is a function $x = \psi(y)$ that can be derived as follows. We know that for all continuous functions $h$ on $[0,1]$, \[
\frac{1}{n} \sum_{k=1}^{n} h(x_k) \to \frac{1}{I} \int_{0}^{1} h(t) \sqrt{f'(t)} \, dt \quad \text{as} \quad n \to \infty.
\]
Here again $I$ is the normalizing constant $\int_{0}^{1} \sqrt{f'(t)} \, dt$. So letting $t = \psi(s)$, use the change of variables formula for integrals to write
\[
\frac{1}{I} \int_{0}^{1} h(t) \sqrt{f'(t)} \, dt = \frac{1}{I} \int_{0}^{1} h(\psi(s)) \sqrt{f'(\psi(s))} \psi'(s) \, ds.
\]
Thus, to transform the $x_k$ into something close to uniform distribution, we would want to have $\frac{d}{ds}(F \circ \psi)(s) = \frac{1}{I} \sqrt{f'(\psi(s))} \psi'(s) = 1$ for all $s$ where $F(t)$ is any antiderivative of $\frac{1}{I} \sqrt{f'(t)}$. This is the same as saying $F \circ \psi(s) = s$ for all $s$ and so $\psi = F^{-1}$.

**Remark 2.3.** Finding $F$ explicitly could be difficult in general. But if $f(x) = x^d$, then you can easily compute that $F(t) = t^{(d+2)/2}$. Hence, $\psi(y) = y^{2/(d+2)}$. So we are expecting that the partition points $x^*_k$ for the optimal partition $P^*_n$ for $f(x) = x^d$ will be very close to $\psi(k/(n+1)) = (k/(n+1))^{2/(d+2)}$. For example, if $f(x) = x^d$, then we expect $x^*_k$ to be approximately $(k/(n+1))^{1/2}$. One can easily check that this is not the precise value, but it is close. It would be worthwhile to know what the value of the error is. There are a number of ways to measure this, but here is one that may be interesting and worthwhile.

What is the asymptotic value of $\sum_{k=1}^{n} (x^*_k - (k/(n+1))^2$.

**Acknowledgements:** The Illinois Geometry Lab team that worked on optimal partitions in Spring, 2012 consisted of Yitian Chen, Yiwang Chen, Ilkyoo Choi, and Michael Lowney. They are the ones who invented the technique in Remark 2.2.

**References**


**J. Rosenblatt**

Department of Mathematics
University of Illinois
Champaign-Urbana, Illinois, USA
email: rosnbltt@illinois.edu