MCTP - Summer Program 2010 (sponsored by NSF)
Minicourse: FOURIER ANALYSIS AND WAVELETS
Instructor: Cristina Pereyra
University of New Mexico
Project #2: **The power of Averaging or Summability Methods**

We have seen that if a 2π-periodic function f is just a little better than continuous, for instance if f has two continuous derivative then the Fourier series of f converges uniformly on $[-\pi, \pi)$, in this project we will see that in this case the partial Fourier sums converge (uniformly) to f. One derivative suffices to guarantee uniform convergence, but the argument is more complicated, one can keep on weakening the assumptions and getting at each stage an even more technical theorem. But if the hope is to show that the partial Fourier sums converge pointwise for continuous functions (forget about uniformly), that dream was shattered in 1892 by a German mathematician Du Bois Reymond. He constructed a CONTINUOUS function whose partial Fourier sums DIVERGED at the origin. In 1926 the Russian mathematician Kolmogorov found an even more dramatic example of an INTEGRABLE function whose partial Fourier sums DIVERGED EVERYWHERE. It was not until 1966 that the Swedish mathematician L. Carleson, one of today's greatest analysts, was able to show that the partial Fourier sums will converge almost everywhere for square integrable functions and in particular for continuous functions (convergence occurs except possibly on a set of measure zero).
However, in an attempt to obtain a convergence result for all continuous functions, mathematicians devised various averaging or *summability methods*. These methods require only knowledge of the Fourier coefficients in order to recover a continuous function as the sum of an appropriate trigonometric series. Along the way, a number of very important approximation techniques in analysis were developed, in particular *convolutions*, and *approximations of the identity*, also known as *good kernels*.
We will start with the same observation (according to T. Körner) that Fejér did when he was 19 years old: "If a sequence is not terribly well behaved, its behavior maybe improved by considering averages. [...] (Any reader discouraged by Fejér's precocity should note that a few years earlier his school considered him so weak in mathematics as to require extra tuition)." T.Körner, *Fourier Analysis*, Cambridge University Press, 1988.
In the second part of this project we describe several kernels that arise naturally in the theory of Fourier series. Some of them (the Fejér and Poisson[1] kernels) are good kernels that generate approximations of the identity; another (the Dirichlet[2] kernel) is equally important but not good in this sense. Finally we learn about smoothing properties of convolutions and approximations of the identity. In particular we can deduce from all the work done in the project that the Fourier series of f with two continuous derivatives converge uniformly to f. As a by-product we get the

[1]Named after the Hungarian mathematician Lipót Fejér (1880-1959) and the French mathematician Simeón Denis Poisson (1781–1840).
[2]Named after the German mathematician Johann Peter Gustav Lejeune Dirichlet (1805-1859).

celebrated Weierstrass' Theorem, and we conclude that the trigonometric functions form a *complete orthonormal system* or an *orthonormal basis*.

This is a long project, it consists of two sub-projects: averaging methods, and good kernels and approximations of the identity. There are additional "homeworks" for further exploration on: Summability matrices, convolution as a smoothing operation, convolution vs Fourier transform, and an introduction to the geometry of $L^2(\mathbb{T})$ that will be the starting point for the third project, when we will go from Fourier series to the finite or discrete Fourier transform.

For more on summability methods and matrices see the handout Section 1.4.2 in the book by Mark Pinsky *Introduction to Fourier Analysis and Wavelets*, the Brooks/Cole Series in Advanced Mathematics, 2002. He also has a lovely discussion of so-called Tauberian theorems where convergence of the averages plus some extra knowledge on the sequence guarantees convergence of the sequence (as opposed to Abelian theorems where convergence of the sequence implies convergence of the averages).

One can deduce from Weiertrass' theorem the famous Weyl's equidistribution theorem, which if we had time we will discuss. If you are curious you can start your search in the handouts from Tom Körner's *Fourier Analysis* Chapter I.3, Cambridge Press, or from Mark Pinsky's *Introduction to Fourier Analysis and Wavelets*, Section 1.4.4 p.57. There have been some recent spectacular results about distributions of primes by Terry Tao (Field medalist 2006) and Greene using harmonic analysis techniques, but this is part of another story.

You might want to explore whether the Gibbs phenomenon is tamed down by summability methods or not. In the Gibbs' Project you were encouraged to tame down the Gibbs Phenomenon in 2(i), the idea there was to consider integral averages of $S_N f(x)$ over the interval $[x - \frac{2\pi}{N}, x + \frac{2\pi}{N}]$ centered at x, and shrinking to $\{x\}$ as N grows to infinity.

Handouts:

- From Pinski's book, Section 1.4.2 *Summability matrices*, Section 1.4.3 *Fejér Means of a Fourier Series*, Section 1.4.4 *Equidistribution Modulo One*, Section 1.4.5 *Hardy's Tauberian Theorem*.

- From Körner's book, Chapter 1 *Introduction*, Chapter 2 *Proof of Fejér's Theorem*, Chapter 3 *Weyl's Equidistribution Theorem*.

- From Elena Prestini's book, Section 3.3. *Tables of Fourier transforms*, Section 3.4. *The Dirac delta function and related topics*, Section 3.5. *Convolution*, Section 3.6. *Filters, noise, and false alarms*, Section 4.12. *Convolution and some special effects.*

Part I: Averaging Methods

Here we explore how different "averaging" methods do not disturb convergence properties of sequences, in fact they could improve them.

1. **Cesàro means:** The first averaging method is named after the Italian mathematician Ernesto Cesàro (1859-1906).

   (a) (Cesàro means) Given a sequence of real numbers $\{s_n\}_{n=1}^{\infty}$ convergent to the number s, that is $\lim_{n\to\infty} s_n = s$. Show that the new sequence of averages or arithmetic means, σ_n, converges and to the same limit.

   $$\sigma_n := \frac{s_1 + s_2 + \cdots + s_n}{n} \to s, \qquad \text{as } n \to \infty$$

   Experiment first with some examples: $s_n = 1$, $s_n = n$, $s_n = 1/n$, etc. Can you find an example of a sequence $\{s_n\}$ so that the sequences of averages converges but not the original sequence?

   (b) Just for fun, if I give you any fraction $\frac{p}{q}$ between zero and one, Can you find a sequence of 1s and 0s so that the averages converge to $\frac{p}{q}$? Notice that such sequence will necessarily have infinitely many 1s and 0s. What if I give you any real number $0 < \alpha < 1$, not necessarily a fraction?

   Note: If we are given a convergent sequence $\{s_n\}$ and we can show the sequence of averages converges to s then the sequence itself has no other choice than to converge to s.

   Looking ahead, we showed in Project #1 that if f is periodic and has two continuous derivatives then its Fourier series converges uniformly. <u>If we can show that the averages of the partial Fourier sums of f, $S_N f$, converge (uniformly) to f,</u> then by the note above, the partial Fourier sums will have no other choice than to converge to f as advertized. The underlined statement is exactly what we will attempt to do in Part II of this project.

2. **Abel means:** The second averaging method is named after the Norwegian mathematician Niels Henrik Abel (1802-1829).

   (a) (Abel means) Given a sequence of real numbers $\{s_n\}_{n=1}^{\infty}$ convergent to the number s, that is $\lim_{n\to\infty} s_n = s$. Show that for $0 \le r < 1$

   $$\lim_{r\to 1^-} (1-r) \sum_{n=0}^{\infty} r^n s_n = s. \tag{1}$$

   Can you find an example of a sequence $\{s_n\}$ so that the limit in (1) exists but the sequence does not converge?

   (b) Notice that we can think of the right hand side of (1) as a "limit of a limit", that is,

   $$\lim_{r\to 1^-} \lim_{N\to\infty} (1-r) \sum_{n=0}^{N} r^n s_n,$$

   what happens if you interchange the limits?

   (c) In order to work with Abel means, it is useful to note the following transformation formula, for $0 \le r < 1$, and for $s_n = a_0 + a_1 + \cdots + a_n$, $n = 0, 1, \ldots,$

   $$(1-r) \sum_{n=0}^{\infty} r^n s_n = \sum_{n=0}^{\infty} r^n a_n.$$

   This identity allows us to go back and forth between a sequence $\{a_n\}$ and its partial sums $\{s_n\}$. Verify the identity.

   Note: If we are given a convergent sequence $\{s_n\}$, and the limit in (1) exists then the sequence itself has no other choice than to converge to s.

3. **Cesàro vs Abel**

   (a) What happens to the sequences of 1s and 0s constructed in 1(b) when you computed the limit in (1)? Is this a coincidence?

   (b) Can you find a sequence $\{s_n\}$ so that the sequence of averages diverges, but the limit in (1) exists? With the information so far gathered, if you wanted to rank our summability methods, which one will be stronger, Abel or Cesàro?

   Hint: Consider the sequences $\{(-1)^n(n+1)\}_{n\ge 0}$....

Homework: Summability matrices

This is a discrete analogue of an approximation of the identity that we will explore in Part II.
First a definition, a *summability matrix* is a doubly infinite array of real numbers $A = \{a_{mn}\}$ defined for $m, n \geq 0$ such that,

- $\lim_{m\to\infty} a_{mn} = 0$ for each $n = 0, 1, 2, \ldots$.
- $\sum_{n=0}^{\infty} a_{mn} = 1$ for each $m = 0, 1, 2, \ldots$.
- $\sum_{n=0}^{\infty} |a_{mn}| \leq C$ for some constant $C > 0$ and for all $m = 0, 1, 2, \ldots$.

A summability matrix defines a linear transformation on the space of bounded sequences. More precisely, let $S = \{s_n\}_{n\geq 0}$ be a bounded sequence, that is there is a a constant $M > 0$ such that $|s_n| \leq M$ for all $n \geq 0$. Define a new sequence $A(S) = \{A_m(S)\}_{m\geq 0}$ by

$$S \to A(S), \qquad A_m(S) = \sum_{n=0}^{\infty} a_{mn}s_n.$$

Make sure that the new sequence is well-defined, that is for each m, the term $A_m(S)$ is finite, moreover the new sequence is a bounded sequence as well.

1. What is the simplest summability matrix you can consider?

2. Can we think of the Cesàro and Abel means as part of the framework of summability matrices?

3. Show that if the sequence $S = \{s_n\}_{n\geq 0}$ converges to the real number s, and A defines a summability matrix, then the new sequence $A(S) = \{A_m(S)\}_{m\geq 0}$ also converges to the real number s.

4. It is natural to compare methods of summability. We say that method A is *stronger* than method C if the matrix A can be factored in the form $A = BC$ where B is another summability matrix. Why on earth would this definition make any sense? Now, with this definition, try to show that Abel is stronger than Cesàro (in fact stronger than any power of Cesàro).

Pinski's handout Sec 1.4.2 may come in handy for this homework.

PART II: GOOD AND NOT SO GOOD KERNELS

1. **Dirichlet Kernel:** Here we explore the Dirichlet kernel that appears very naturally when rewriting the partial Fourier sums as a convolution.

   (a) Write the Nth partial Fourier sum of f as an integral of f times a certain function,

   $$S_N f(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) D_N(\theta - y)\, dy =: f * D_N(\theta) \quad \text{(convolution)} \qquad (2)$$

   This function D_N is called the *Dirichlet kernel*. Notice that D_N does not depend on f. Who is the function D_N?

   (b) Show that

   $$D_N(\theta) = \sum_{k=-N}^{N} e^{ik\theta} = 1 + 2\sum_{k=1}^{N} \cos(k\theta) = \frac{\sin\left(\frac{(2N+1)\theta}{2}\right)}{\sin\frac{\theta}{2}}.$$

   (c) What is the maximum value of D_N? Sketch graphs of D_N for different values of N, if need be and you know how, with MATLAB. Note that $D_N(\theta)$ does NOT converge to zero as $N \to \infty$ for any value of θ, why?

   (d) Show that the Dirichlet kernels have mean value one on $[-\pi, \pi]$.

   (e) However the mean value of the absolute values grows like $\ln N$, how can this be possible? More precisely, show that there is a number $C > 0$ independent of N such that

   $$L_N := \frac{1}{2\pi} \int_{-\pi}^{\pi} |D_N(y)|\, dy \geq C \ln N.$$

   The numbers L_N are called the Lebesgue numbers. if you attempted problem 3(i) in Gibbs' Project you already encountered this calculation.

2. **Fejér Kernel:** Averaging the Dirichlet kernels we get the better Fejér kernels, and instead of the partial Fourier sums the better Cesàro sums. Here we explore basic properties of these objects.

   Recall that D_N is a trigonometric polynomial of degree N, with coefficients equal to 1 for $-N \leq n \leq N$, and 0 for all other values of n. The Dirichlet kernel D_N is 2π-periodic. See Part II: 1(b) in this project for different formulas for the Dirichlet kernel.

   (a) Let the Fejér kernels be the averages of the Dirichlet kernels,
   $$F_N(\theta) := \frac{D_0(\theta) + D_1(\theta) + \cdots + D_{N-1}(\theta)}{N}.$$
   Show that
   $$F_N(\theta) = \frac{1}{N}\left[\frac{\sin(N\theta/2)}{\sin(\theta/2)}\right]^2.$$

   (b) What is the maximum value of F_N?

   (c) Show that the Fejér kernels are positive, and have mean value one on $[-\pi, \pi]$. Show that $F_N \to 0$ uniformly as $n \to \infty$ for $\delta \leq |\theta| \leq \pi$. Moreover show that for each $\delta > 0$,
   $$\lim_{N\to\infty} \int_{\delta\leq|\theta|\leq\pi} F_N(\theta)\, d\theta = 0.$$
   that is, the mass is accumulating at the origin. Sketch graphs of F_N for different values of N, if need be with MATLAB.

   (d) Show that if $\sigma_N(f)$ denotes the average of the partial Fourier sums, that is
   $$\sigma_N(f) := \frac{S_0(f) + S_1(f) + \cdots + S_{N-1}(f)}{N}.$$
   Then $\sigma_N(f) = f * F_N$, the Cesàro sums.

   (e) Interchanging the order of summation, and counting the number of appearances of each Fourier summand $\widehat{f}(k)e^{ik\theta}$ gives a representation of $\sigma_N f$ as a weighted average of the Fourier coefficients of f corresponding to frequencies $|k| \leq N$, where the coefficients corresponding to smaller frequencies are weighted most heavily. Can you be more precise?

   (f) What is $\widehat{\sigma_N(f)}(n)$? This is an example of a *Fourier multiplier*, why do you think that name was chosen? Another example of a Fourier multiplier is the partial Fourier sum $S_N f$, why? (look at $\widehat{S_N(f)}(n)$). Can you draw a picture comparing the multipliers?

3. **Poisson Kernel:** Using Abel means of the Dirichlet kernels we stumble upon the Poisson kernel, and the Abel sums replace the partial Fourier sums.

The *Poisson kernel* $P_r(\theta)$ is defined for $r \in [0,1)$ by

$$P_r(\theta) := (1-r)\sum_{N=0}^{\infty} r^N D_N(\theta).$$

Notice that the Poisson kernel is indexed by all real numbers r between 0 and 1, not by the discrete positive integers N as in the Dirichlet and Fejér kernels. We are interested now in the behavior as r increases towards 1, instead of $N \to \infty$. If you have encountered the Poisson kernel in a complex analysis or a PDE class, you might not recognize it in this guise.

(a) Verify that the Poisson kernel can be expressed as follows,

$$P_r(\theta) = \sum_{n=-\infty}^{\infty} r^{|n|} e^{in\theta} = \frac{1-r^2}{1-2r\cos\theta + r^2}.$$

Deduce now that the Poisson kernel is non-negative (for $r \in [0,1)$). These representations should be more familiar to some of you. The Poisson kernel is the prototype of a very special class of functions defined on the unit disc $\{z \in \mathbb{C} : |z| \leq 1\}$, can you name it?

(b) What is the maximum value of P_r?

(c) Show that the Poisson kernels have mean value one on $[-\pi,\pi]$. Show that $P_r \to 0$ uniformly as $r \to 1^-$ for $\delta \leq |\theta| \leq \pi$. Moreover show that for each $\delta > 0$,

$$\lim_{r\to 1^-} \int_{\delta \leq |\theta| \leq \pi} P_r(\theta)\, d\theta = 0.$$

that is, the mass is accumulating at the origin. Sketch a graph of P_r for different values of r, if need be with MATLAB.

(d) Let $A_r(f)$ denote the Abel mean of the partial Fourier sums, that is $A_r(f) := (1-r)\sum_{N=0}^{\infty} r^N S_N(f)$. Show that for $r \in [0,1)$,

$$A_r(f)(\theta) := \sum_{n=-\infty}^{\infty} r^{|n|}\widehat{f}(n) e^{in\theta}.$$

Thus the Abel mean is formed from f by multiplying the Fourier coefficients $\widehat{f}(n)$ by the corresponding "damping" factor $r^{|n|}$. What is $\widehat{A_r(f)}(n)$? this is another example of a Fourier multiplier, see 2(f).

(e) The Abel mean arises from the Poisson kernel by convolution, just as the Cesàro mean arises from the Fejér kernel by convolution. Verify that for integrable functions $f : \mathbb{T} \to \mathbb{C}$,

$$A_r(f)(\theta) = (P_r * f)(\theta).$$

4. **Good Kernels:** A family $\{K_n\}_{n=1}^{\infty}$ of real-valued integrable functions on the circle, $K_n : [-\pi, \pi) \to \mathbb{R}$, is a *family of good kernels* if it satisfies these three properties:

   - The K_n all have mean value 1.
   - The integrals of their absolute values are uniformly bounded. That is there is an $M > 0$ so that for ALL n, $\int_{-\pi}^{\pi} |K_n(\theta)|\, d\theta \leq M$.
   - Mass is concentrated at the origin. More precisely, for each $\delta > 0$,

   $$\int_{\delta \leq |\theta| \leq \pi} |K_n(\theta)|\, d\theta \to 0 \qquad \text{as } n \to \infty.$$

     This is achieved if $K_n \to 0$ uniformly as $n \to \infty$ for $\delta \leq |\theta| \leq \pi$.

   (a) We already have a couple of examples of good kernels. Which ones? We also have an example of a kernel that is not so good, which one?

   (b) Suppose you had to construct a family of good kernels, what would be the simplest possible example you can think of? Check that $K_n(x) = 2\pi n \chi_{[-1/2n, 1/2n]}(x)$ is a good kernel (Draw some pictures you do NOT need MATLAB!). Compute $K_n * f(x)$. Suppose f is continuous, what is $\lim_{n\to\infty} K_n * f(x)$? Prove it (you may want to use the Fundamental Theorem of Calculus). This is an instance of the Lebesgue Differentiation Theorem that says that the averages of an integrable function over intervals shrinking to a point converge to the value of the function at the point almost everywhere.

5. Good kernels are also called *approximations of the identity* for there is a theorem that says that the convolutions of these kernels K_n with a reasonable function f converge, as $n \to \infty$, to the function f, at least at points of continuity. Moreover, a priori, $f * K_n$ is at the very least continuous (by the smoothing properties of convolution). For those of you that have done advanced calculus, can the convergence be uniform in general? The second part of the theorem says that we can approximate continuous functions f UNIFORMLY by $K_n * f$, if K_n is a good kernel. Try to prove this theorem. (Körner's handout Chapter 2 has the proof).

   With all this information, can we approximate continuous functions uniformly by trigonometric polynomials? If yes, you have proved the celebrated **Weierstrass' Theorem:** *trigonometric polynomials are dense in the continuous functions with respect to the uniform norm on compact sets.*

6. Fejér's theorem says that the Cesàro means $\sigma_N f$ of a continuous function converge uniformly to f. We showed that if f is 2π-periodic and twice continuously differentiable ($f \in C^2(\mathbb{T})$) then the partial Fourier sums $S_n f$ converge uniformly (but we did not know to which function). Show that if $f \in C^2(\mathbb{T})$ then the partial Fourier sums $S_N f$ converge uniformly to f as $N \to \infty$. Hint: Part I - Exercise 1(a), and the note will help.

Homework: Convolution is a Smoothing operation

1. We have learned that convolution appears naturally in the context of Fourier series. In this part we want to argue in favor of the statement: *convolution is a smoothing operation.*

   (a) Let f be the 2π-periodic function that on the interval $[-\pi, \pi)$ coincides with one when $x \in [0, 1]$ and zero otherwise. Calculate $f * f$ explicitly (and draw a picture of it). What happens to the "support" of the outcome? You started with two discontinuous functions, what happened when you convolved them? Did you get a differentiable function? Calculate know $f * f * f$, can you see any improvement?

   This is an example of how *convolution improves smoothness*. Even more is true: if the convolved functions are smooth then the convolution absorbs the smoothness from each of them, as the following exercise illustrates.

   (b) Suppose f, $g : [-\pi, \pi) \to \mathbb{C}$, f is k-times continuously differentiable ($f \in C^k(\mathbb{T})$) , and g is m-times continuously differentiable ($g \in C^m(\mathbb{T})$). Show that $f * g$ is $(k+m)$-times differentiable ($f * g \in C^{k+m}(\mathbb{T})$). Furthermore the following formula holds:

   $$(f * g)^{(k+m)} = f^{(k)} * g^{(m)}.$$

   (It suffices to assume that the functions $f^{(k)}$ and $g^{(m)}$ are bounded and integrable to conclude that $f * g \in C^{k+m}(\mathbb{T})$.) Hint: try induction!

   (c) Another instance of how convolution keeps "the best features" from each function is the following exercise.

   Show that if f is integrable and P is a trigonometric polynomial of order N then so is $f * P$.

   If we could find a kernel that is a trigonometric polynomial, then $f * K_n$ will be itself a trigonometric polynomial. Have we found such kernel already?

2. We explore here the interplay between convolution and products under Fourier transformations.

   (a) A very important feature of the Fourier theory is that convolutions are transformed into products, and viceversa. Verify that for reasonable functions

   $$\widehat{f * g}(n) = \widehat{f}(n)\widehat{g}(n).$$

   (b) With this information at hand and the work done in Part II, what are the Fourier coefficients of the Dirichlet, the Fejér, and the Poisson kernels?

Homework: Excursion into $L^2(\mathbb{T})$

Define the space of square integrable functions $L^2(\mathbb{T})$ to consist of those functions f such that

$$\int_{-\pi}^{\pi} |f(\theta)|^2 d\theta < \infty.$$

This is a normed space with norm

$$\|f\|_{L^2(\mathbb{T})} := \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} |f(\theta)|^2 d\theta \right)^{1/2}.$$

In fact if the method of integration used is Lebesgue integration, then this is a *complete normed space* or *Banach space* in the sense that Cauchy sequences are convergent to square integrable functions.

1. Show that the partial Fourier sums of $f \in C^2(\mathbb{T})$ converge to f in $L^2(\mathbb{T})$, i.e.

$$\lim_{N\to\infty} \int_{-\pi}^{\pi} |S_N f(\theta) - f(\theta)|^2 d\theta = 0.$$

2. A "density" argument allows one to pass from continuous functions to square integrable functions. The density result that you need is *continuously twice differentiable functions are dense in* $L^2(\mathbb{T})$. Can you sketch the argument? You will need an extra piece of information to close the argument, for those of you familiar with the language in functional analysis: *S_N are uniformly (in N) bounded linear operators in* $L^2(\mathbb{T})$, namely there exists a constant $C > 0$ (in fact $C = 1$ will work) such that for all $g \in L^2(\mathbb{T})$, and for all $N \in \mathbb{N}$,

$$\|S_N g\|_{L^2(\mathbb{T})} \le C \|g\|_{L^2(\mathbb{T})}.$$

3. What is all this fuzz with square integrable functions? It has a lot of additional geometric structure encoded in the word *Hilbert space*. There is an inner product, defined for all $f, g \in L^2(\mathbb{T})$,

$$\langle f, g \rangle := \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta)\overline{g(\theta)}\, d\theta.$$

With this inner product you have a notion of orthogonality: f is orthogonal to g iff $\langle f, g \rangle = 0$. Verify that the trigonometric functions $\{e_n(\theta) := e^{in\theta}\}_{n\in\mathbb{Z}}$ form an orthonormal system, that is:

$$\langle e_n, e_m \rangle = \delta(n-m) = \begin{cases} 1 & \text{if } n = m \\ 0 & \text{otherwise} \end{cases}.$$

All together we have argüed that the trigonometric functions form an *orthonormal basis in* $L^2(\mathbb{T})$. In particular, the trigonometric polynomial $S_N f$ is the *best approximation to f in $L^2(\mathbb{T})$ among all trigonometric polynomials of degree smaller or equal than N*, in fact it is the *orthogonal projection of f onto the subspace of all trigonometric polynomials of degree smaller or equal than N.*