LECTURE 9

N-DIMENSIONAL SPACE

When I became a professor, after 30 years of active research at Bell Telephone Laboratories, mainly in the math research department, I recalled that professors are supposed to think and digest past experiences. So I put my feet up on the desk and began to consider my past. In the early years I had been mainly in computing so naturally I was involved in many large projects that required computing. Thinking about how things worked out on several of the large engineering systems I was partially involved in, I began, now that I had some distance from them, to see that they had some common elements. Slowly I began to realize that the design problems all took place in a space of n-dimensions, where n is the number of independent parameters. Yes, we build 3 dimensional objects, but their design is in a high dimensional space, one dimension for each design parameter.

I also need high dimensional spaces so that later proofs will become intuitively obvious to you without filling in the details rigorously. Hence we will discuss n-dimensional space now.

You think you live in three dimensions, but in many respects you live in a two dimensional space. For example, in the random walk of life, if you meet a person you then have a reasonable chance of meeting that person again. But in a world of three dimensions you do not! Consider the fish in the sea who potentially live in three dimensions. They go along the surface, or on the bottom, reducing things to two dimensions, or they go in schools, or they assemble at one place at the same time, such as a river mouth, a beach, the Sargasso sea, etc. They cannot expect to find a mate if they wander the open ocean in three dimensions. Again, if you want airplanes to hit each other, you assemble them near an airport, put them in two dimensional levels of flight, or send them in a group; truly random flight would have fewer accidents than we now have!

N-dimensional space is a mathematical construct which we must investigate if we are to understand what happens to us when we wander there during a design problem. In two dimensions we have Pythagoras' theorem that for a right triangle the square of the hypotenuse equals the sum of the squares of the other two sides. In three dimensions we ask for the length of the diagonal of a rectangular block, Figure 9-1. To find it we first draw a diagonal on one face, apply Pythagoras' theorem, and then take that as one side with the other side the third dimension, which is at right angles, and again from the Pythagorean theorem we get that the square of the diagonal is the sum of the squares of the three perpendicular sides. It is obvious from this proof, and the necessary symmetry of the formula, that as you go to higher and higher dimensions you will still have the square of the diagonal as the sum of the squares of the individual mutually perpendicular sides,

$$D^2 = \text{SUM}[i=1,n;\ x_i^2]$$

where the x_i are the lengths of the sides of the rectangular block in n-dimensions.

Continuing with the geometric approach, planes in the space will be simply linear combinations of the x_i, and a sphere about a point will be all points that are at the fixed distance (the radius) from the given point.

We need the volume of the n-dimensional sphere to get an idea of the size of a piece of restricted space. But first we need the Stirling approximation for n!, which I will derive so that you will see most of the details and be convinced that what is coming later is true, rather than on hearsay.

A product like n! is hard to handle, so we take the log of n! which becomes

$$\ln n! = \text{SUM}[k=1,n;\ \ln k]$$

where, of course, the ln is the logarithm to the base e. Sums remind us that they are related to integrals, so we start with the integral

$$\text{INT}[1,n;\ \ln x\ dx]$$

We apply integration by parts, (since we recognize that the ln x arose from integrating an algebraic function and hence it will be removed in the next step). Pick U = ln x, dV = dx, then

$$\text{INT}[1,n;\ \ln x\ dx] = \{x \ln x - x\}\big|1,n$$

$$= n \ln n - n + 1$$

On the other hand, if we apply the trapezoid rule to the integral of ln x we will get, Figure 9-2,

$$\text{INT}[1,n;\ln x\ dx] \sim (1/2)\ln 1 + \ln 2 + \ln 3 + \ \ldots\ + (1/2)\ln n$$

Since ln 1 = 0, adding (1/2)ln n to both terms we get, finally,

$$\text{SUM}[k=1,n;\ \ln k] \sim n \ln n - n + 1 + (1/2)\ln n$$

Undo the logs by taking the exponential of both sides

$$n! \sim Cn^n e^{-n} (n)^{1/2}$$

where C is some constant (not far from e) independent of n, since we are approximating an integral by the trapezoid rule and the error in the trapezoid approximation increases more and more slowly as n grows larger and larger, and C is the limiting value. This is the first form of Stirling's formula. We will not waste time to deriving the limiting, at infinity, value of the constant C which turns out to be $\sqrt{(2\pi)} = 2.5066\ldots$ ($e = 2.71828\ldots$). Thus we finally have the usual Stirling's formula for the factorial

$$n! \sim n^n e^{-n} \sqrt{(2\pi n)}$$

The following table shows the quality of the Stirling approximation to n!

n	Stirling	True	Stirling/true
1	0.92214	1	0.92214
2	1.91900	2	0.95950
3	5.83621	6	0.97270
4	23.50518	24	0.97942
5	118.01916	120	0.98349
6	710.07818	720	0.98622
7	4,980.3958	5,040	0.98817
8	39,902.3958	40,320	0.98964
9	359,536.87	362,880	0.99079
10	3,598,695.6	3,628,800	0.99170

Note that as the numbers get larger and larger the ratio approaches 1 but the differences get greater and greater! If you consider the two functions

$$f(n) = n + \sqrt{n}$$

$$g(n) = n$$

then the limit of the ratio $f(n)/g(n)$, as n approaches infinity, is 1, but as in the table the difference

$$f(n) - g(n) = \sqrt{n}$$

grows larger and larger as n increases.

We need to extend the factorial function to all positive real numbers, hence we introduce the <u>gamma function</u> in the form of an integral

$$\Gamma(n) = \text{INT}[0,\infty;\ x^{n-1}e^{-x}\,dx]$$

which converges for all $n > 0$. For $n > 1$ we again integrate by parts, this time using the $dV = e^{-x}\,dx$ and the $U = x^{n-1}$. At the two limits the integrated part is zero, and we have the reduction formula

$$\Gamma(n) = (n-1)\Gamma(n-1)$$

with $\Gamma(1) = 1$

Thus the gamma function takes on the values $(n-1)!$ at the positive integers n, and it provides a natural way of extending the factorial to all positive numbers since the integral exist whenever $n > 0$.

We will need

$$\Gamma(1/2) = \text{INT}[0,\infty;\ x^{-1/2}e^{-x}\ dx]$$

Set $x = t^2$, hence $dx = 2t\ dt$, and we have (using symmetry in the last step)

$$\Gamma(1/2) = 2\text{INT}[0,\infty;\exp(-t^2\ dt] = \text{INT}[-\infty,\infty;\ \exp(-t^2)\ dt]$$

We now use a standard trick to evaluate this integral. We take the product of two of the integrals, one with x and one with y as their variables.

$$\Gamma^2(1/2) = \text{INT}[0,\infty;\ \text{INT}[0,\infty\varepsilon\ \exp-(x^2 + y^2)\ dx]\ dy]$$

The $x^2 + y^2$ suggests polar coordinates, so we convert

$$= \text{INT}[0,2\pi;\ \text{INT}[0,\infty;\ \exp(-r^2)\ r\ dr\ d\theta]$$

The angle integration is easy, the exponential is now also easy, and we get, finally,

$$\Gamma^2(1/2) = \pi$$

Thus

$$\Gamma(1/2) = \sqrt{\pi} = 1.77245...$$

We now turn to the volume of an n-dimensional sphere (or hypersphere if you wish). Clearly the volume of a cube in n dimensions and of side x is x^n. A little reflection and you will believe that the formula for the volume of an n-dimensional sphere must have the form

$$\text{Volume} = C_n r^n$$

where C_n is a suitable constant. In the case $n = 2$ the constant is π, in the case $n = 1$, it is 2 (when you think about it). In three dimensions we have $C_3 = 4\pi/3$.

We start with same trick as we used for the gamma function of 1/2, except that this time we take the product of n of the integrals, each with a different x_i. Thinking of the volume of a sphere we see that it is the sum of shells, and each element of the sum has a volume which is the corresponding shell area multiplied by the thickness, dr. For a sphere the value for the surface area can be obtained by differentiating the volume of the sphere with respect to the radius r,

$$\text{Surface} = dV_n(r)/dr = nC_n r^{n-1}$$

and hence the elements of volume are

$$\{dV_n(r)/dr\}dr = nC_n r^{n-1} dr$$

We have, therefore, on setting $r^2 = t$

$$\Gamma^n(1/2) = \pi^{n/2} = \mathrm{INT}[0,\infty;\ \exp(-r^2)\ \{dV_n(r)/dr\}dr]$$

$$= (nC_n/2)\mathrm{INT}[0,\infty; e^{-t}\ t^{(n/2-1)}\ dt]$$

$$= (nC_n/2)\Gamma(n/2) = C_n\ \Gamma(n/2 + 1)$$

from which we get

$$C_n = \pi^{n/2}/\Gamma(n/2 + 1)$$

It is easy to see that

$$C_n = [(2\pi)/n]C_{n-2}$$

and we can compute the following table.

dimension n	coefficient C_n	
1	2	= 2.00000...
2	π	= 3.14159...
3	$4\pi/3$	= 4.11879...
4	$\pi^2/2$	= 4.93480...
5	$8\pi^2/15$	= 5.26379...
6	$\pi^3/6$	= 5.16771...
7	$16\pi^3/105$	= 4.72477...
8	$\pi^4/24$	= 4.05871...
9	$32\pi^4/945$	= 3.29850...
10	$\pi^5/120$	= 2.55010...
2k	$\pi^k/k!$	--> 0

Thus we see that the coefficient C_n increases up to n = 5 and then decreases towards 0. For spheres of unit radius this means that the volume of the sphere approaches 0 as n increases. If the radius is r, then we have for the volume, and using n = 2k for convenience (since the actual numbers vary smoothly as n increases and the odd dimensional spaces are messier to compute),

$$C_n r^n = (\pi r^2)^k/k! \ \text{-->}\ 0 \quad \text{as } k \ \text{-->}\ \infty$$

No matter how large the radius, r, increasing the number of dimensions, n, will ultimately produce a sphere of arbitrarily small volume.

Next we look at the relative amount of the volume close to the surface of a n-dimensional sphere. Let the radius of the sphere be r, and the inner radius of the shell be $r(1 - \epsilon)$, then the <u>relative volume</u> of the shell is

$$[C_n r^n - C_n r^n (1 - \epsilon)^n]/C_n r^n = 1 - (1 - \epsilon)^n$$

For large n, no matter how thin the shell is (relative to the radius), almost all the volume is in the shell and there is almost nothing inside. As we say, the volume is almost all on the surface. Even in 3 dimensions the unit sphere has 7/8 ths of its volume within 1/2 of the surface. In n-dimensions there is $1 - 1/2^n$ within 1/2 of the radius from the surface.

This has importance in design; it means that almost surely the optimal design will be on the surface and will not be inside as you might think from taking the calculus and doing optimizations in that course. The calculus methods are usually inappropriate for finding the optimum in high dimensional spaces. This is not strange at all; generally speaking the best design is pushing one or more of the parameters to their extreme - obviously you are on the surface of the feasible region of design!

Next we turn to looking at the diagonal of an n-dimensional cube, say the vector from the origin to the point (1,1,...,1). The cosine of the angle between this line and any axis is given by definition as the ratio of the component along the axis, which is clearly 1, to the length of the line which is $\sqrt{n}$. Hence

$$\cos\theta = 1/\sqrt{n} \;\text{-->}\; 0 \quad \text{and} \quad \theta \;\text{-->}\; \pi/2$$

Therefore, for large n the diagonal is almost perpendicular to every coordinate axis!

If we use the points with coordinates (± 1, ± 1, ... ± 1) then there are 2^n such diagonal lines which are all almost perpendicular to the coordinate axes. For n = 10, for example, this amounts to 1024 such almost perpendicular lines.

I need the angle between two lines, and while you may remember that it is the vector dot product, I propose to derive it again to bring more understanding about what is going on. [Aside; I have found it very valuable in important situations to review all the basic derivations that are involved so I have a firm feeling for what is going on.] Take two points x and y with their corresponding coordinates x_i and y_i, Figure 9-3. Then applying the law of cosines in the plane of the three points x, y, and the origin we have

$$C^2 = X^2 + Y^2 - 2XY \cos\theta$$

where X and Y are the lengths of the lines to the points x and y. But the C comes from using the differences of the coordinates in each direction

$$C^2 = \text{SUM}[k=1,n;\ (x_k - y_k)^2 = X^2 + Y^2 - 2\text{SUM}[k=1,n;\ x_k y_k]$$

Comparing the two expressions we see that

$$\cos\theta = \text{SUM}[k=1,n;\ x_k y_k]/XY$$

We now apply this formula to two lines drawn from the origin to random points of the form

$$(\pm 1, \pm 1, \ldots, \pm 1)$$

The dot product of these factors, taken at random, is again random ± 1's and these are to be added n times, while the length of each is again $\sqrt{n}$, hence (note the n in the denominator)

$$\cos\theta = \text{SUM}[k=1,n; (\pm 1)]/n$$

and by <u>the weak law of large numbers</u> this approaches 0 for increasing n, <u>almost surely</u>. But there are 2^n different such random vectors, and given any one fixed vector then any other of these 2^n random vectors is <u>almost surely almost perpendicular</u> to it! N-dimensions is indeed vast!

In linear algebra and other courses you learned to find the set of perpendicular axes and then represent everything in terms of these coordinates, but you see that in n-dimensions there are, after you find the n mutually perpendicular coordinate directions, 2^n other directions that are <u>almost perpendicular</u> to those you have found! The theory and practice of linear algebra are quite different!

Lastly, to further convince you that your intuitions about high dimensional spaces are not very good, I will produce another paradox which I will need in later Lectures. We begin with a 4x4 square and divide it into 4 unit squares in each of which we draw a unit circle, Figure 9-4. Next we draw a circle about the center of the square with radius just touching the four circles on their insides. Its radius must be, from the Figure 9-4,

$$r_2 = \sqrt{2} - 1 = 0.414\ldots$$

Now in three dimensions you will have a 4x4x4 cube, and 8 spheres of unit radius. The inner sphere will touch each outer sphere along the line to their center will have a radius of

$$r_3 = \sqrt{3} - 1 = 0.732\ldots$$

Think of why this must be larger than for two dimensions.

Going to n dimensions, you have a 4x4x...x4 cube, and 2^n spheres, one in each of the corners, and with each touching its n adjacent neighbors. The inner sphere, touching on the inside all of the spheres, will have a radius of

$$r_n = \sqrt{n} - 1$$

Examine this carefully! Are you sure of it? If not, why not? Where will you object to the reasoning?

Once satisfied that it is correct we apply it to the case of n = 10 dimensions. You have for the radius of the inner sphere

$$r_{10} = \sqrt{10} - 1 > 2$$

and in 10 dimensions the inner sphere reaches outside the surrounding cube! Yes, the sphere is convex, yes it touches each of the 1024 packed spheres on the inside, yet it reaches outside the cube!

So much for your raw intuition about n-dimensional space, but remember that the n-dimensional space is where the design of complex objects generally takes place. You had better get an improved feeling for n-dimensional space by thinking about the things just presented, until you begin to see how they can be true, indeed why they must be true. Else you will be in trouble the next time you get into a complex design problem. Perhaps you should calculate the radii of the various dimensions, as well as go back to the angles between the diagonals and the axes, and see how it is that it can happen.

It is now necessary to note carefully, that I have done all this in the classical Euclidean space using the Pythagorean distance where the sum of squares of the differences of the coordinates is the distance between the points squared. Mathematicians call this distance L_2.

The space L_1 uses not the sum of the squares, but rather the sum of the distances, much as you must do in traveling in a city with a rectangular grid of streets. It is the sum of the differences between the two locations that tells you how far you must go. In the computing field this is often called the "Hamming distance" for reasons that will appear in a later Lecture. In this space a circle in 2 dimensions looks like a square standing on a point, Figure 9-5. In three dimensions it is like a cube standing on a point, etc. Now you can better see how it is that in the circle paradox above the inner sphere can get outside the cube.

There is a third, commonly used, metric, (they are all metrics = distance functions), called L_∞, or <u>Chebyshev distance</u>. Here we have that the distance is the maximum coordinate difference, regardless of any other differences, Figure 9-6. In this space a circle is a square, a 3 dimensional sphere is a cube, and you see that in this case the inner circle in the circle paradox has 0 radius in all dimensions.

These are all examples of a <u>metric</u>, a measure of distance. The conventional conditions on a metric $D(x,y)$ between two points x and y are:

1. $D(x,y) \geq 0$ (non negative)

2. $D(x,y) = 0$ if and only if $x = y$ (Identity)

3. $D(x,y) = D(y,x)$ (symmetry)

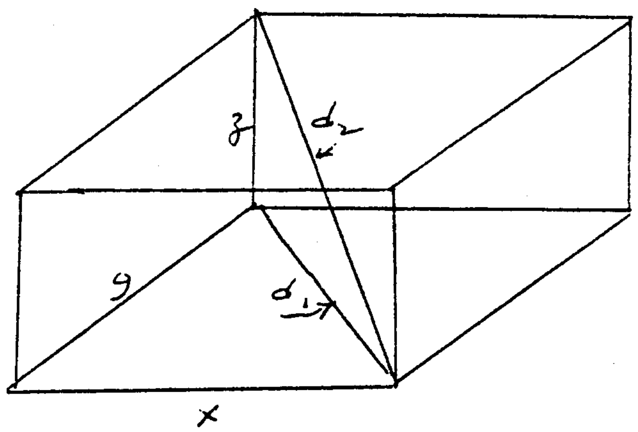
4. $D(x,y) + D(y,z) \geq D(x,z)$ (triangle inequality)

It is left to you to verify that the three metrics, L_∞, L_2 and L_1, (Chebyshev, Pythagoras, and Hamming), all satisfy these conditions.

The truth is, in complex design, for various coordinates we may use any of the three metrics, all mixed up together, so the design space is not as portrayed above, but is a mess of bits and pieces. The L_2 metric is connected with least squares, obviously, and the other two, L_∞ and L_1, are more like comparisons. In making comparisons in real life, you generally use either the maximum difference, L_∞, in any one trait as sufficient to distinguish two things, or sometimes, as in strings of bits, it is the number of differences that matters, and the sum of the squares does not enter, hence the L_1 distance is used. This is increasingly true, for example, in pattern identification in AI.

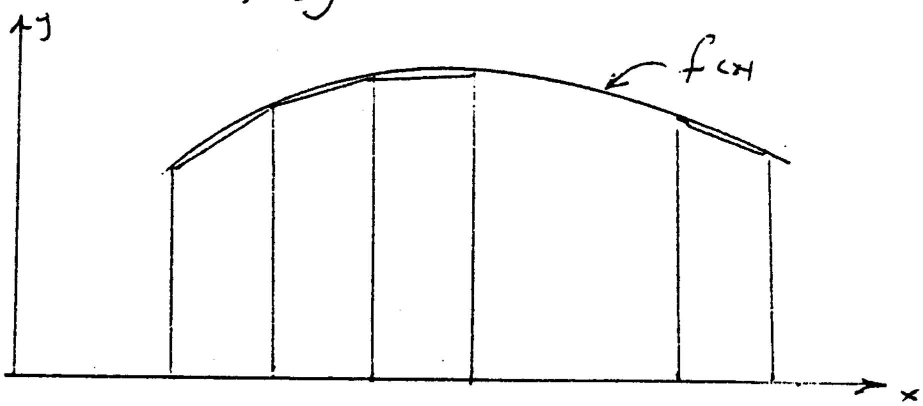
Unfortunately, the above is all too true, and it is seldom pointed out to you. They never told me a thing about it! I will need many of the results in later Lectures, but in general, after this exposure, you should be better prepared than you were for complex design and for carefully examining the space in which the design occurs, as I have tried to do here. Messy as it is, fundamentally it is where the design occurs and where you must search for an acceptable design.

Since L_1 and L_∞ are not familiar let me expand the remarks on the three metrics. L_2 is the natural distance function to use in physical and geometric situations including the data reduction from physical measurements. Thus you find least squares, L_2, throughout physics. But when the subject matter is intellectual judgments then the other two distance functions are generally preferable, and this is slowly coming into use, though we still find the Chi square test, which is obviously a measure for L_2, used widely when some other suitable test should be used.



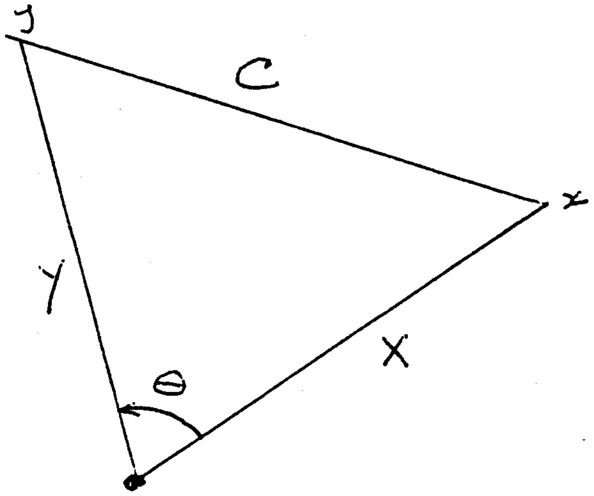
Pythagorean distance

Figure 9-1



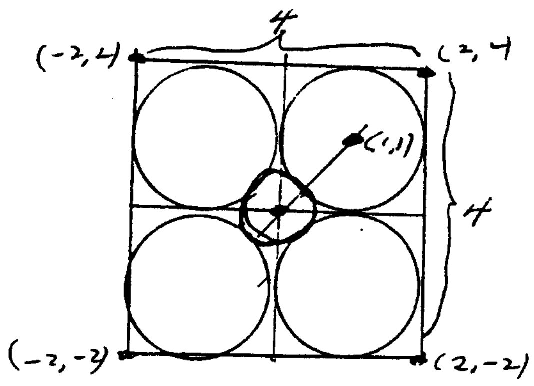
Trapazoid Rule

Figure 9-2



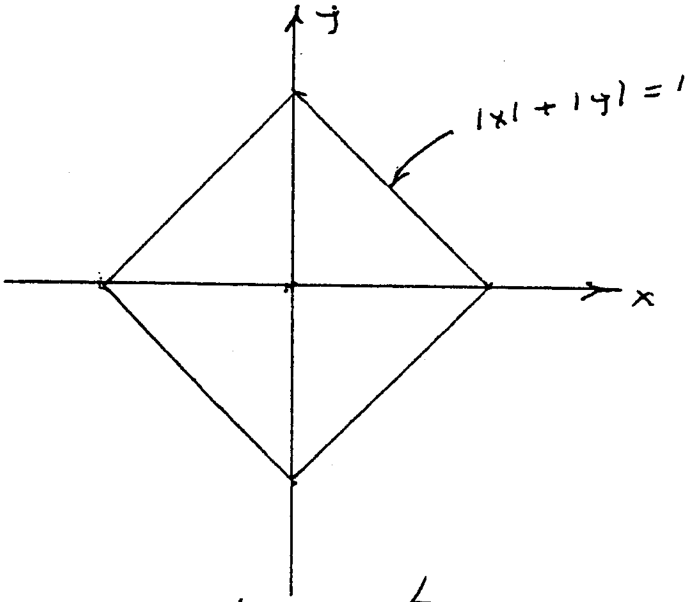
Law of Cosines

Figure 9-3



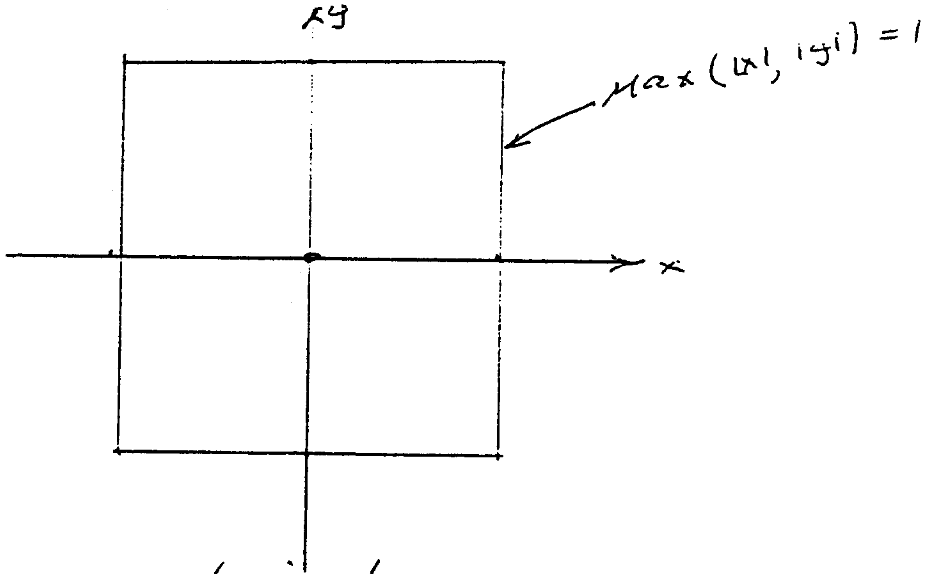
Four Circle Paradox

Figure 9-4



Circle in L_1

Figure 9-5



circle in L_∞

Figure 9-6