

MA 323 Geometric Modelling

Course Notes: Day 04

Data Fitting and Interpolation

David L. Finn

December 3rd, 2004

Today, we consider two related problems; the interpolation problem and the data fitting problem. The interpolation problem takes the form:

Given a collection of data points and a model construction method, determine the “controls” (geometric primitives) for the model so that the model contains all the data points.

All the model construction methods that we have considered so far are designed to solve interpolation problems. Each data point will belong to the curve produced so far. However, not every model construction problem will be designed to solve an interpolation problem. For such methods, solving the interpolation problem is a method for constructing a preliminary model. The controls of the model can then be “tweaked” to change the model, possibly without changing the interpolation of the data points.

The data fitting problem is a similar problem with the exception that instead of having the model pass through every data point, the model needs to pass close to all the data points. The interpolation problem and the data fitting problem are similar because they both use the same information, a collection of data points and a model construction problem. In addition, the result and methodology are similar; determine the “controls” of the model.

4.1 The Interpolation Problem

As stated above, the interpolation problem is to construct a model that passes through a set of data points. Typically, the data points are organized in such a manner that determines the relative location of the points in constructing the model. If no organization is done, it is possible then to obtain very different models from the same set of data points. Think of a connect the dot puzzle in which the same dots are given in different orders. Radically different pictures can then result. Typically, the organization of the data needs to be done first. In practice, the data is obtained through a sampling or scanning method, and thus has a natural order from the data collection method.

The interpolation problem is not necessarily a purely geometric problem. It is frequently posed as a parametric problem with the parameter normally being time or distance. For several of the later methods, parameters are important in the model construction algorithm, and thus must be given to construct the model. For the model construction methods, we

have considered so far, a pure geometric solution still applies but it does not specify when you are at each point, which may be an important part of the model depending on the exact application. For instance, suppose the geometric modelling problem involves specifying the location of a robot navigating a maze. In this problem, one may want to specify the location of the robot in the maze at specific times. Thus, the problem is a temporal problem in addition to a geometric problem.

The parametric interpolation problem for curves is typically stated as follows:

Given a sequence of points p_i and corresponding times t_i find a parametric curve $c(t)$ such that $c(t_i) = p_i$.

One may still solve this problem as a pure geometric problem ignoring the times. One then only has to determine a parameterization of the model curve that gives the location p_i at the correct times. We will use this method in the discussion of the interpolation problem for piecewise linear and piecewise circular curves below.

4.2 Parametric Interpolation with Piecewise Linear Curves

Recall that a line between points A and B when the positions of A and B are described by vectors from a fixed origin (i.e. A and B are described in the usual coordinate axes in either the xy plane or xyz space), the line segment between A and B may be described parametrically by

$$L(\tau) = (1 - \tau)A + \tau B.$$

Notice $L(0) = A$ and $L(1) = B$. If we want to pass through A at a different time say t_0 and B at a different time say t_1 (we will assume that $t_0 < t_1$), we define the line by using composition with linear function τ from the interval $[t_0, t_1]$ to $[0, 1]$, i.e. $\tau(t_0) = 0$ and $\tau(t_1) = 1$.

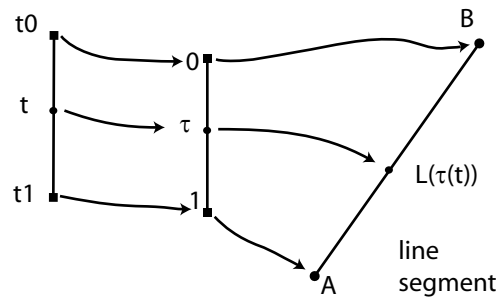


Figure 1: Functional Representation of Parametric Description of Line

The desired linear function $\tau(t)$ is

$$\tau(t) = \frac{t - t_0}{t_1 - t_0}.$$

Therefore, by composition, we can define a parametric line passing through A at t_0 and B at t_1 by $L(\tau(t))$ given by

$$\begin{aligned} L(t) &= \left(1 - \frac{t - t_0}{t_1 - t_0}\right) A + \frac{t - t_0}{t_1 - t_0} B \\ &= \frac{t_1 - t}{t_1 - t_0} A + \frac{t - t_0}{t_1 - t_0} B \end{aligned}$$

A piecewise linear curve satisfying $c(t_i) = p_i$ is then defined as a piecewise parametric function as

$$c(t) = \begin{cases} \frac{t_1-t}{t_1-t_0} p_0 + \frac{t-t_0}{t_1-t_0} p_1 & \text{for } t_0 \leq t < t_1 \\ \frac{t_2-t}{t_2-t_1} p_1 + \frac{t-t_1}{t_2-t_1} p_2 & \text{for } t_1 \leq t < t_2 \\ \vdots & \vdots \\ \frac{t_n-t}{t_n-t_{n-1}} p_{n-1} + \frac{t-t_{n-1}}{t_n-t_{n-1}} p_n & \text{for } t_{n-1} \leq t \leq t_n \end{cases}$$

This function is defined purely on the interval $t_0 \leq t \leq t_n$, as we have no data outside of this region. To solve outside the region in an extrapolation problem, and other methods need to be applied. This type of problem does not often occur in geometric modelling, so we will not consider it.

4.3 Parametric Interpolation with Piecewise Circular Curves

To solve the parametric interpolation problem with piecewise circular curves, one method is to solve the problem geometrically, and then choose the parameterization of the curve to fit the temporal data. This is similar to the solution for interpolation using piecewise linear curves. An obvious choice for the parameterization of a circle is the trigonometric parameterization,

$$c(t) = \begin{bmatrix} h + r \cos(\omega(t)) \\ k + r \sin(\omega(t)) \end{bmatrix}$$

where r is the radius of the circle, $[h, k]$ is the center of the circle, and $\omega(t)$ is a function representing the parameterization of the circle which ensures that $c(t_i) = p_i$. The values of r , h , and k are determined geometrically as in the example below given points p_0 , p_1 , p_2 . The function $\omega(t)$ can be found by first solving for an angle θ_i so that

$$p_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} h + r \cos(\theta_i) \\ k + r \sin(\theta_i) \end{bmatrix}.$$

This can be accomplished by solving $(y_i - k)/(x_i - h) = \tan(\theta_i)$ and adjusting for the quadrant. Once, we have the angles θ_i (up to the addition of a multiple of 2π), the function $\omega(t)$ can be found by piecewise linear interpolation on each circle; that is fitting a line to each consecutive pair (t_i, θ_i) .

Example: Find a parameterization $c(t)$ of the (smooth) circular curve through $p_0 = (0, 1)$, $p_1 = (1, 1)$, $p_2 = (2, -1)$, $p_3 = (3, 1)$ with $c(0) = (0, 1)$, $c(1) = (1, 1)$, $c(2) = (2, -1)$, and $c(4) = (3, 1)$. We shall use the first three points to define the first arc, and the last two to define the second arc. As per the geometric construction, we define $l_1(s_1) = (\frac{1}{2}, 1 - s_1)$ and $l_2(s_2) = (\frac{3}{2} - 2s_2, -s_2)$ as the perpendicular bisectors of p_0p_1 , p_1p_2 to find the first circular arc. Then, solving for the point of intersection of l_1 and l_2 , we have the center of the circle as $(h, k) = (\frac{1}{2}, -\frac{1}{2})$, and the radius as $r = \frac{1}{2}\sqrt{10} = 1.5811$. Thus, we need to solve for $\theta_0, \theta_1, \theta_2$

$$\begin{aligned} [0, 1] &= \left[\frac{1}{2} + \frac{1}{2}\sqrt{10} \cos(\theta_0), -\frac{1}{2} + \frac{1}{2}\sqrt{10} \sin(\theta_0) \right] \\ [1, 1] &= \left[\frac{1}{2} + \frac{1}{2}\sqrt{10} \cos(\theta_1), -\frac{1}{2} + \frac{1}{2}\sqrt{10} \sin(\theta_1) \right] \\ [2, -1] &= \left[\frac{1}{2} + \frac{1}{2}\sqrt{10} \cos(\theta_2), -\frac{1}{2} + \frac{1}{2}\sqrt{10} \sin(\theta_2) \right]. \end{aligned}$$

Solving, we get $\theta_0 = \tan^{-1}(-3) + \pi = 1.8925$, $\theta_1 = \tan^{-1}(3) = 1.2490$, $\theta_2 = \tan^{-1}(-1/3) = -0.3218$. Therefore, fitting a line to $\omega(0) = \theta_0$ and $\omega(1) = \theta_1$ and a second line to $\omega(1) = \theta_1$ and $\omega(2) = \theta_2$, we have

$$\begin{aligned}\omega_1(t) &= 1.8925 - 0.6435t \text{ for } 0 \leq t \leq 1 \\ \omega_2(t) &= 1.2490 - 1.5708(t - 1) \text{ for } 1 \leq t \leq 2\end{aligned}$$

For the second circular arc, we have the radial line for the first arc at $(2, -1)$ as $l_1(s_1) = [2 + \frac{3}{2}s_1, -1 - \frac{1}{2}s_1]$ and the perpendicular bisector of p_2p_3 as $l_2(s_2) = [4 + 2s_2, -4s_2]$. The center of the circle is at the intersection of these two lines, and thus is $(h, k) = (\frac{5}{2}, -2)$ and the radius of the circle is $r = \sqrt{10} = 3.1623$. We then need to solve

$$\begin{aligned}[2, -1] &= [5 + \sqrt{10} \cos(\theta_3), -2 + \sqrt{10} \sin(\theta_3)] \\ [6, 1] &= [5 + \sqrt{10} \cos(\theta_4), -2 + \sqrt{10} \sin(\theta_4)].\end{aligned}$$

These equations yield $\theta_3 = \tan^{-1}(-1/3) + \pi = 2.8198$ and $\theta_4 = \tan^{-1}(3) = 1.2490$, but this yields a nonsmooth circular curve (we have a cusp in the curve at $[2, -1]$). To rectify this situation, add a full rotation of the circle to θ_4 so that $\theta_4 = \tan^{-1}(3) + 2\pi = 7.5322$. Thus, fitting a line to $\omega(2) = 2.8198$ and $\omega(4) = 7.5322$, we have

$$\omega_3(t) = 2.8198 + 2.3562(t - 2) \text{ for } 2 \leq t \leq 4.$$

The circular curve thus is

$$c(t) = \begin{cases} [\frac{1}{2} + \frac{1}{2}\sqrt{10} \cos(\omega_1(t)), -\frac{1}{2} - \frac{1}{2}\sqrt{10} \sin(\omega_1(t))] & \text{for } 0 \leq t \leq 1 \\ [\frac{1}{2} + \frac{1}{2}\sqrt{10} \cos(\omega_2(t)), -\frac{1}{2} - \frac{1}{2}\sqrt{10} \sin(\omega_2(t))] & \text{for } 1 \leq t \leq 2 \\ [5 + \sqrt{10} \cos(\omega_3(t)), -2 + \sqrt{10} \sin(\omega_3(t))] & \text{for } 2 \leq t \leq 4. \end{cases}$$

see diagrams below, where the diagram on the left represents choosing the nonsmooth solution above and the diagram on the right represents the smooth solution.

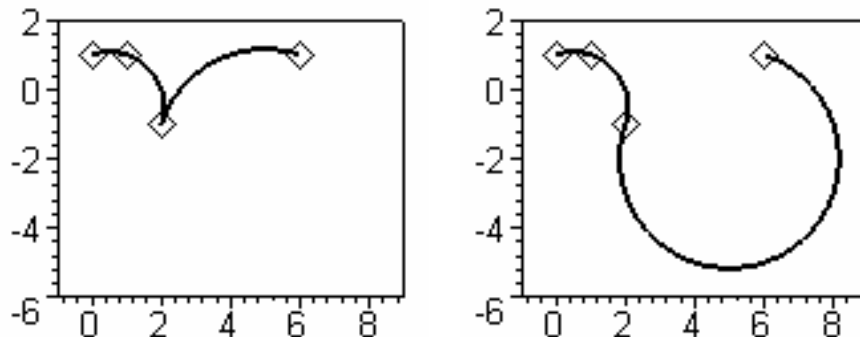


Figure 2: Solving Interpolation Problem with Piecewise Circular Curves

The trigonometric parameterization has a small inconvenience. One has to use inverse trigonometric functions, and worry which angle to use when solving, so that we interpolate

the correct points and the resulting curve is smooth. This is an inconvenience. There are other parameterizations of a circle that allow a pure algebraic solution (only involve solving a quadratic equation) and eliminate any checking. Specifically, we parameterize a circle with center $[h, k]$ and radius r as a rational function

$$c(t) = \begin{bmatrix} h + r \frac{1-\xi(t)^2}{1+\xi(t)^2} \\ k + r \frac{2\xi(t)}{1+\xi(t)^2} \end{bmatrix}.$$

Note that $(1 - \xi^2)^2 + (2\xi)^2 = 1 - 2\xi^2 + \xi^4 + 4\xi^2 = 1 + 2\xi^2 + \xi^4 = (1 + \xi^2)^2$, so this is a parameterization of a circle of radius r with center $[h, k]$.

In this parameterization, the full circle is described with $\xi \in]-\infty, \infty[$ with each value corresponding to a different point on the circle, and the limiting value $\xi \rightarrow \pm\infty$ corresponding to the point $[h - r, k]$ while $\xi = 0$ corresponds to $[h + r, k]$. This is completely different from the trigonometric parameterization of the circle, where the full circle is described by a closed interval $[0, 2\pi]$. It is useful to change the position corresponding to $\xi = 0$ and more importantly $\xi = \pm\infty$, by rotating the coordinate system. This is important to ensure that the arc created does not need to pass through $\xi = \pm\infty$. This can be accomplished by introducing a rotation parameter ϕ so that

$$c(t) = \left[h + r \frac{\cos(\phi)(1-\xi(t)^2) - 2\sin(\phi)\xi(t)}{1+\xi(t)^2}, k + r \frac{\sin(\phi)(1-\xi(t)^2) + 2\cos(\phi)\xi(t)}{1+\xi(t)^2} \right].$$

We shall see that there is a more elegant formulation for piecewise circular arcs which requires use of projective geometry that is more amenable to solving interpolation problems.

4.4 Data Fitting - an approximation problem

A problem that is similar to the interpolation problem is to find a model that is close to collection of data points, but does not exactly pass through them. Such problems arise frequently because when dealing with sampled data, the data is frequently *noisy*. This means that a solution to an interpolation problem does not usually provide a good model, as it may need to have a high degree of variation to pass through all the points, see diagrams below. The *noise* (slight measurement errors) in the data can greatly affect the solution to an interpolation problem, and cause the curve to be widely different. Rather than use interpolation, an alternate method of constructing a model from sampled data points is to construct a model that is close to the data points. Unlike in the solution to the interpolation problem, the model does not have to pass through the given data points. Typically, the model minimizes a certain objective function with respect to all possible models of a given type.

To describe a data fitting problem more concretely, consider the case where one samples 20 points from a circle of radius 2 centered at the point (1,2) and stores the data with two decimal places with slight error in the measurements (see table below and diagram below). If one uses an interpolation method to solve the problem with either a piecewise linear curve or a piecewise circular curve (depending on the chosen tangent line in the approximation one can get a different but similar curve) see figures below. This may not be an acceptable model, especially if one knows or suspects that the data comes from a circle.

$p_1 = (3.14, 2.15)$	$p_{11} = (-1.00, 2.04)$
$p_2 = (2.98, 2.78)$	$p_{12} = (-0.79, 1.57)$
$p_3 = (2.67, 3.28)$	$p_{13} = (-0.45, 0.92)$
$p_4 = (2.25, 3.71)$	$p_{14} = (-0.17, 0.44)$
$p_5 = (1.66, 3.98)$	$p_{15} = (0.43, 0.22)$
$p_6 = (1.17, 4.07)$	$p_{16} = (1.04, 0.06)$
$p_7 = (0.41, 4.09)$	$p_{17} = (1.79, 0.26)$
$p_8 = (-0.06, 3.69)$	$p_{18} = (2.25, 0.43)$
$p_9 = (-0.60, 3.20)$	$p_{19} = (2.72, 0.99)$
$p_{10} = (-0.83, 2.74)$	$p_{20} = (3.09, 1.42)$

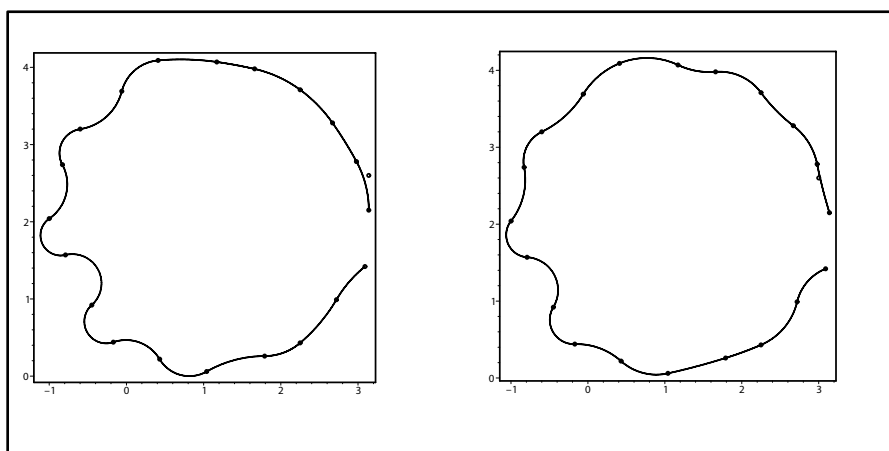


Figure 3: Data and Piecewise Circular Curves

Suppose it is known that the data does come from a circle. One would like to recover the exact circle at least approximately that passes through these points. A piecewise circular curve should do this, if the data is 100% accurate which will not happen in practice. We then want to find the circle that comes closest to passing through the data points. This is an approximation problem: given a set of data points, find an object that is close to all the data points. In solving this problem, one typically has an idea from examining the data (performing statistical tests either analytically using statistics, just looking at the data). In our example, from looking at the data you might suspect that there is a circle or one can know from the application that there is an underlying circle. First note a circle has three degrees of freedom; two for the center of the circle and one for the radius. Using the general equation of a circle $(x - h)^2 + (y - k)^2 = r^2$ in expanded form, this can be rearranged as

$$x^2 + y^2 = 2xh + 2yk + r^2 - h^2 - k^2 \quad \text{or} \quad x^2 + y^2 = ax + by + c$$

where $a = 2h$, $b = 2k$, $c = r^2 - h^2 - k^2$. From the data given in table (?), we get 20 linear equations as we know the points (x, y) for each data point and we are looking for the parameters a , b , c .

Unless the data is perfect, we will not be able to find three parameters that solve all 20 equations. Therefore, we set up a minimization problem, we look for the parameters a , b , c that minimize the *total distance* function

$$f(a, b, c) = \sum_{i=1}^{20} (x_i^2 + y_i^2 - ax_i - by_i - c)^2.$$

Taking the partial derivatives, and setting equal to zero (finding the critical points of f) yields three equations and three unknowns. In fact, because the function $f(a, b, c)$ is a quadratic function in the variables a, b, c , finding the critical points yields three linear equations. For the data indicated above, we find that $a = 2.19, b = 4.20, c = -1.55$, which yields a circle with center $o = [1.1, 2.1]$ and radius $r = 2.02$. This is close to the values of the circle from which the data was generated, but not exactly the same.

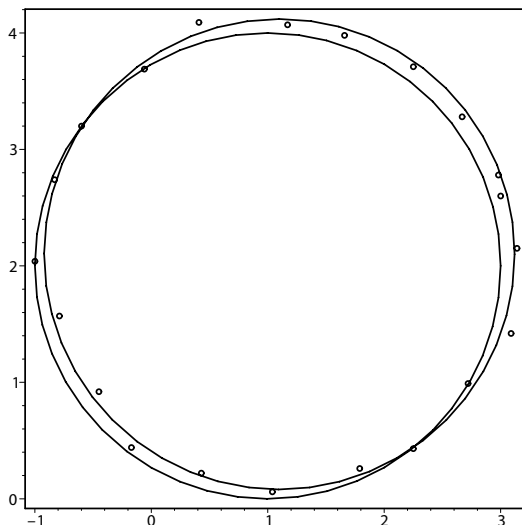


Figure 4: Fitted and True Circle to Data Points

One can use other objective functions. The *total squared distance function* is the standard objective function, as it acts like the sum of the distances. It is not the actual distance function. One can define the objective function using the actual distance between the model curve and the data. This objective function is then harder to compute, but it is possible.

4.5 Method of Least Squares - matrix approach

The method used above is called the method of least squares. This method can be applied as above using calculus. However, it has another formulation in terms of matrices and linear algebra which is more commonly used. The reason for this is one may view least squares as an attempt to solve an overdetermined system of equations. We desire an approximate solution x to

$$Ax = b$$

where A is a matrix with more rows than columns. This may also be viewed as projection problem; project the vector b onto the column space of the matrix A (the vector space obtained as linear combination of the columns of the matrix A), and then solve the corresponding linear system.

To explain the projection view of least squares, let us consider the system of equations in vector form

$$\mathbf{v}_1 x_1 + \mathbf{v}_2 x_2 = \mathbf{b},$$

where $\mathbf{v}_1, \mathbf{v}_2, \mathbf{b}$ are three dimensional vectors. Generically, such a system is not solvable since there are three equations and only two unknowns. To obtain an approximate solution,

we can project the vector \mathbf{b} onto the plane spanned by \mathbf{v}_1 and \mathbf{v}_2 and solve the associated system, that is we write $\mathbf{b} = \mathbf{p} + \mathbf{q}$ where \mathbf{p} is the projection onto the plane and \mathbf{q} is normal to plane (orthogonal to \mathbf{v}_1 and \mathbf{v}_2). The projection onto the plane spanned by $\mathbf{v}_1, \mathbf{v}_2$ can be found by taking the dot product the vector vector equation with each vector \mathbf{v}_1 and \mathbf{v}_2 to get the system of equations (the normal equations)

$$\begin{aligned}(\mathbf{v}_1 \cdot \mathbf{v}_1) x_1 + (\mathbf{v}_1 \cdot \mathbf{v}_2) x_2 &= (\mathbf{v}_1 \cdot \mathbf{b}) \\(\mathbf{v}_2 \cdot \mathbf{v}_1) x_1 + (\mathbf{v}_2 \cdot \mathbf{v}_2) x_2 &= (\mathbf{v}_2 \cdot \mathbf{b}),\end{aligned}$$

noting that if $\mathbf{b} = \mathbf{p} + \mathbf{q}$ then $\mathbf{v}_i \cdot \mathbf{b} = \mathbf{v}_i \cdot \mathbf{p}$. The above system is solvable (as long as the vectors \mathbf{v}_1 and \mathbf{v}_2 are linearly independent), with solution

$$\begin{aligned}x_1 &= \frac{(\mathbf{v}_2 \cdot \mathbf{v}_2)(\mathbf{v}_1 \cdot \mathbf{b}) - (\mathbf{v}_1 \cdot \mathbf{v}_2)(\mathbf{v}_2 \cdot \mathbf{b})}{(\mathbf{v}_1 \cdot \mathbf{v}_1)(\mathbf{v}_2 \cdot \mathbf{v}_2) - (\mathbf{v}_1 \cdot \mathbf{v}_2)^2} \\x_2 &= \frac{-(\mathbf{v}_2 \cdot \mathbf{v}_1)(\mathbf{v}_1 \cdot \mathbf{b}) + (\mathbf{v}_2 \cdot \mathbf{v}_2)(\mathbf{v}_2 \cdot \mathbf{b})}{(\mathbf{v}_1 \cdot \mathbf{v}_1)(\mathbf{v}_2 \cdot \mathbf{v}_2) - (\mathbf{v}_1 \cdot \mathbf{v}_2)^2}.\end{aligned}$$

The projection of \mathbf{b} onto the plane spanned by \mathbf{v}_1 and \mathbf{v}_2 is then

$$\begin{aligned}\mathbf{p} &= \frac{(\mathbf{v}_2 \cdot \mathbf{v}_2)(\mathbf{v}_1 \cdot \mathbf{b}) - (\mathbf{v}_1 \cdot \mathbf{v}_2)(\mathbf{v}_2 \cdot \mathbf{b})}{(\mathbf{v}_1 \cdot \mathbf{v}_1)(\mathbf{v}_2 \cdot \mathbf{v}_2) - (\mathbf{v}_1 \cdot \mathbf{v}_2)^2} \mathbf{v}_1 + \\&\quad \frac{-(\mathbf{v}_2 \cdot \mathbf{v}_1)(\mathbf{v}_1 \cdot \mathbf{b}) + (\mathbf{v}_2 \cdot \mathbf{v}_2)(\mathbf{v}_2 \cdot \mathbf{b})}{(\mathbf{v}_1 \cdot \mathbf{v}_1)(\mathbf{v}_2 \cdot \mathbf{v}_2) - (\mathbf{v}_1 \cdot \mathbf{v}_2)^2} \mathbf{v}_2.\end{aligned}$$

Instead of working with vectors (or points), the method of least squares is normally written in matrix form, as the number of equations is typically quite large (i.e. the twenty data points in the circle example can be considered as small for a typical example).

To cast the method of least squares in matrix form, first note that the matrix form of $\mathbf{v}_1 x_1 + \mathbf{v}_2 x_2 = \mathbf{b}$ is

$$\begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \\ v_{31} & v_{32} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

where \mathbf{v}_i has components v_{1i}, v_{2i}, v_{3i} . Next, we recall that in matrix multiplication

$$AB = C$$

the entry c_{ij} of C is obtained as the dotproduct of the i th row of A with the j th column of B . In the least squares solution, we obtain the normal equations by dotting with the column vectors of the original matrix. This is equivalent to forming a matrix A^T (called the transpose matrix) with each row of A^T a column of A , and then performing matrix multiplication on both sides;

$$\begin{bmatrix} v_{11} & v_{21} & v_{31} \\ v_{12} & v_{22} & v_{32} \end{bmatrix} \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \\ v_{31} & v_{32} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} v_{11} & v_{21} & v_{31} \\ v_{12} & v_{22} & v_{32} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

The solution for x_1 and x_2 is then obtained by inverting the matrix and performing the matrix multiplication to get

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \left(\begin{bmatrix} v_{11} & v_{21} & v_{31} \\ v_{12} & v_{22} & v_{32} \end{bmatrix} \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \\ v_{31} & v_{32} \end{bmatrix} \right)^{-1} \begin{bmatrix} v_{11} & v_{21} & v_{31} \\ v_{12} & v_{22} & v_{32} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

and the projection is formed by performing the matrix multiplication;

$$\begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \\ v_{31} & v_{32} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

using x_1 and x_2 defined above.

In full generality, the method of least squares in matrix form is a method to obtain an approximate solution to an overdetermined system $Ax = b$ (more equation than unknowns). It is accomplished by multiplying both sides by the transpose of the matrix A to get the normal equations

$$A^T A x = A^T b$$

obtaining the least squares solution

$$x = (A^T A)^{-1} A^T b$$

and the projection onto the column space of A

$$\Pi(b) = Ax = A(A^T A)^{-1} A^T b,$$

where Π represents the projection operation.

Typically, one does not invert the matrix $A^T A$ to solve the normal equations, as inverting a matrix is not numerically stable and can yield inaccurate results. One typically applies Gaussian elimination to solve the normal equations and then projects. The subtleties of numerical linear algebra is not the point of this course, but the reader should be aware of the problems and methods used, and where error can arise in calculations.

As an example of the matrix version of least squares, let's consider fitting the five points in the table below to a line $y = mx + b$.

x	0.24	0.37	0.55	0.78	0.94
y	2.05	4.06	6.01	7.99	9.98

We need to solve the system of equations in matrix form

$$\begin{bmatrix} 0.24 & 1 \\ 0.37 & 1 \\ 0.55 & 1 \\ 0.78 & 1 \\ 0.94 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 2.05 \\ 4.06 \\ 6.01 \\ 7.99 \\ 9.98 \end{bmatrix}$$

and the associated normal equations

$$\begin{bmatrix} 1.989 & 2.880 \\ 2.880 & 5.00 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 20.913 \\ 30.090 \end{bmatrix}.$$

The solution is then $m = 10.85$ and $b = -0.23$ which gives the data

x	y data	y fitted
0.24	2.05	2.3740
0.37	4.06	3.7845
0.55	6.01	5.7375
0.78	7.99	8.2330
0.94	9.98	9.9690

4.6 Parametric Data Fitting

A common problem that arises in geometric modelling is the application of data fitting to parametric data, where each data point is associated to a parametric representation of the point on the model curve. The data in this case is usually obtained by some sort of sampling method. For instance, suppose the data in the table below was obtained by sampling from a circle at equal angles of a parametric circle $c(t) = [h + r \cos(2\pi t), k + r \sin(2\pi t)]$

t	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
x	4.75	3.69	2.43	1.46	1.04	1.55	2.41	3.73	4.64	5.07
y	6.32	7.06	7.01	6.26	5.08	3.89	3.29	3.17	3.85	5.12

Applying least squares to the data above to solve for (h, k, r) yields $h = 3.077$, $k = 5.105$, and $r = 2.007$. The data yields 20 equations to be solved are of the form

$$x = h + r \cos(2\pi t) \quad \text{and} \quad y = k + r \sin(2\pi t).$$

The equations are all linear in h , k , and r , so least squares can be applied in matrix form.

The parametric approximation problem solved by least squares is a very useful provided that one chooses an appropriate parametric curve. The choice of a parametric curve will be affected by the data, and will affect the model created. At present, we have only discussed lines and circles which restricts the type of curves that we can approximate. It is easy to imagine other types of parametric curves with parameters that can be fitted to data. The choice of parametric curve that is used to fit the data is largely determined by the application, and some knowledge of the data.

A natural question that one may ask is how to fit data to piecewise curves. Now, fitting data to piecewise curves requires the data is grouped according to the curve segment. Each data point needs to be associated with a curve segment. Unless the data naturally arises with a given curve segment, this will be a hard problem to solve. One will need to analyze the data to attack the problem with the present methods. There are more sophisticated methods for constructing piecewise defined curves that allow more amenable data fitting.

Exercises

- Find a parameterization of the circle that fits the points $(1, 0)$, $(2, 2)$, $(3, 4)$, with $c(0) = (1, 0)$ and $c(1) = (3, 4)$
 - Using the rational parameterization $[h + r \frac{1-\xi^2}{1+\xi^2}, k + r \frac{2\xi}{1+\xi^2}]$ with $\xi = a_0 + a_1 t$.
 - Using the trigonometric parameterization $[h + r \cos(\omega), k + r \sin(\omega)]$ with $\omega = b_0 + b_1 t$.
 - Compare the values for $t = 0.5$ and $t = 0.8$, are they the close or drastically different?
- Given the data below

t	0.0	1.0	2.0	3.5	4.0
x	0.0	1.0	2.0	3.0	3.0
y	0.0	1.0	1.0	2.0	3.0

- Find the piecewise linear curve that interpolates the parametric data

- (b) Find the smooth piecewise circular curve that interpolates the data using the trigonometric parameterization with initial tangent being the y -axis
- (c) Compare the points for $t = 1.5$ and $t = 2.5$ on the above curves. Are they even close together?

3. Apply the method of least squares to solve the system of linear equations

$$\begin{bmatrix} 1 & 2 & 5 \\ 3 & 1 & -1 \\ 2 & -1 & 4 \\ 5 & 1 & 2 \\ 3 & 1 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 3 \\ 4 \end{bmatrix}$$

4. Given the data in the table below

t	0.2	0.5	0.6	0.8	1.2
x	1.0	2.0	3.0	2.5	2.0
y	2.0	3.0	3.0	4.0	5.0

- (a) Find a straight line of the form $[x_0 + a_1 t, y_0 + b_1 t]$ that fits the data using least squares.
- (b) Find a parametric circle of the form $[x_0 + r \cos(t), y_0 + r \sin(t)]$ that best fits the data.
- (c) Find a parabola of the form $[x_0 + a_1 t, y_0 + a_2 t^2]$ that best fits the data.
- (d) Which would you say is the best overall model of the data? Why?