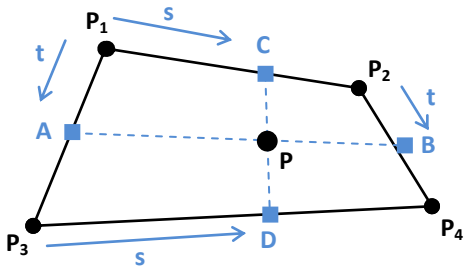
Irregular Bilinear Interpolation

Irregular bilinear interpolation determines the value at a given point by taking the weighted average of its four closest neighbors. This algorithm is nonlinear and more computationally intensive than standard bilinear interpolation. If the data being interpolated is regularly spaced use the standard version detailed here.

The figure below show an irregular point set. Points P represents the new point and points 1-4 are the bounding points. The distance s is the percentage along the vertical sides to point P, the distance t is the horizontal. Points $A - D$ are intermediate points used for calculation.



The edges are represented as parametric lines using the equations:

$$x = x_0 + \alpha \cdot t$$
$$y = y_0 + \beta \cdot t$$

The intermediate points A - D can be described in terms of parametric lines through the corner points P_1 - P_4.

Point A
$$A_x = x_1 + x_{31} \cdot t$$
$$A_y = y_1 + y_{31} \cdot t$$

Point A
$$B_x = x_2 + x_{42} \cdot t$$
$$B_y = y_2 + y_{42} \cdot t$$

Point C
$$C_x = x_1 + x_{21} \cdot s$$
$$C_y = y_1 + y_{21} \cdot s$$

Point D
$$D_x = x_3 + x_{43} \cdot s$$
$$D_y = y_3 + y_{43} \cdot s$$

Where the intermediate variables are defined as:

$$x_{31} = x_3 - x_1$$
$$y_{31} = y_3 - y_1$$
$$x_{42} = x_4 - x_2$$
$$\dots$$

Using points A - D the point P can be solved for by two different parametric lines.

$$P_x = A_x + (B_x - A_x) \cdot s$$
$$P_y = A_y + (B_y - A_y) \cdot s$$

and

$$P_x = C_x + (D_x - C_x) \cdot t$$
$$P_y = C_y + (D_y - C_y) \cdot t$$

The value of P can be attained by solving either pair of these equations, but there are conditions that only one or the other set will give a viable answer.

Method 1

Using the first set of equations the value of s can be formulated by rearranging the equation for y_P.

$$P_y = A_y + (B_y - A_y) \cdot s \qquad \rightarrow \qquad s = \frac{P_y - A_y}{B_y - A_y}$$

The value of s can then be plugged into the equation for x_P.

$$P_x = A_x + (B_x - A_x) \cdot \left(\frac{P_y - A_y}{B_y - A_y}\right)$$

This equation is then rearranged:

$$0 = (B_x - A_x)(P_y - A_y) - (P_x - A_x)(B_y - A_y)$$

The values of A_x, B_x, A_y, and B_y are plugged in to place the solution in terms of t.

$$0 = \big((x_2 + x_{42} \cdot t) - (x_1 + x_{31} \cdot t)\big)\Big(P_y - (y_1 + y_{31} \cdot t)\Big) - \big(P_x - (x_1 + x_{31} \cdot t)\big)\big((y_2 + y_{42} \cdot t) - (y_1 + y_{31} \cdot t)\big)$$

Through a significant bit of algebra this equation can be condensed and factored into the terms of the second order polynomial:

$$At^2 + Bt + C = 0$$

Where the variables A - C are:

$$\begin{aligned} A &= x_{31}y_{42} - y_{31}x_{42} \\ B &= P_y(x_{42} - x_{31}) - P_x(y_{42} - y_{31}) + x_{31}y_2 - y_{31}x_2 + x_1y_{42} - y_1x_{42} \\ C &= P_yx_{21} - P_xy_{21} + x_1y_2 - x_2y_1 \end{aligned}$$

The value of t can be solved for using the quadratic equation:

$$(t_1, t_2) = \frac{-B \pm \sqrt{B^2 - 4AC}}{2C}$$

Since this equation is quadratic, it is likely that two separate roots will be found when solving the equation. When the object of this method is interpolation (not extrapolation), determining which factor is real and which is false is simple. For any interpolation the value of t must be between 0 and 1.

Using the value of t the value of s can be solved for using the value of t.

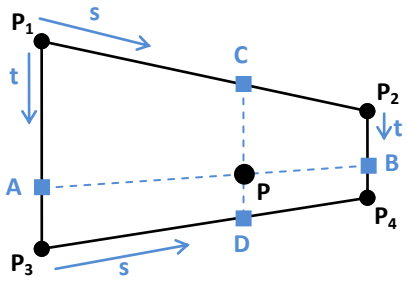
$$s = \frac{P_y - A_y}{B_y - A_y} \qquad \rightarrow \qquad s = \frac{P_y - y_1 - y_{31}t}{y_2 + y_{42}t - y_1 - y_{31}t}$$

After calculating the s and t values, the value at point P is calculated using the standard bilinear interpolation formula.

$$\begin{aligned} P = {} & P_1 \cdot (1 - s) \cdot (1 - t) + \\ & P_2 \cdot (s) \cdot (1 - t) + \\ & P_3 \cdot (1 - s) \cdot (t) + \\ & P_4 \cdot (s) \cdot (t) \end{aligned}$$

Method 2

If the vertical uprights of the interpolation area are parallel, the solution for the quadratic equation will return complex values. If this occurs a second method focusing on the horizontal sections of the viewing area can be used instead.



The same calculations from the first method are used but with the points C and D instead. Following the same calculations, the values of the quadratic coefficients become:

$$A = x_{21}y_{43} - y_{21}x_{43}$$
$$B = P_y(x_{43} - x_{21}) - P_x(y_{43} - y_{21}) + x_1y_{43} - y_1x_{43} + x_{21}y_3 - y_{21}x_3$$
$$C = P_yx_{31} - P_xy_{31} + x_1y_3 - x_3y_1$$

The value of s is then calculated using the quadratic equation.

$$(s_1, s_2) = \frac{-B \pm \sqrt{B^2 - 4AC}}{2C}$$

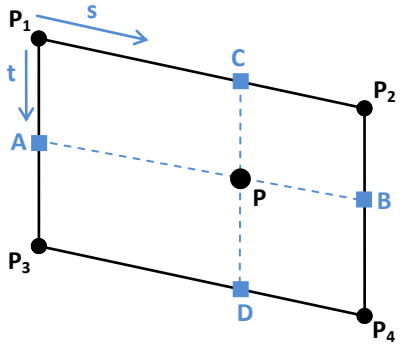
The value of t can be solved for

$$t = \frac{P_y - C_y}{D_y - C_y} \qquad \rightarrow \qquad t = \frac{P_y - y_1 - y_{21} \cdot s}{y_3 + y_{43} \cdot s - y_1 - y_{21} \cdot s}$$

And s and t can be used in the standard bilinear interpolation formula listed above.

Method 3

If the interpolation area becomes a parallelogram, a third method is required to solve for the value at P.



With both pairs of edges parallel, the solution for the point P becomes linear and only marginally more complicated than the standard bilinear interpolation method. Take the original equations for A and B:

$$A_x = x_1 + x_{31} \cdot t \qquad B_x = x_2 + x_{42} \cdot t$$
$$A_y = y_1 + y_{31} \cdot t \qquad B_y = y_2 + y_{42} \cdot t$$

And the solution for P given A and B:

$$P_x = A_x + (B_x - A_x) \cdot s$$
$$P_y = A_y + (B_y - A_y) \cdot s$$

Plugging in the values for A and B gives:

$$P_x = x_1 + x_{31}t + x_2 s + x_{31}st - x_1 s - x_{42}st$$
$$P_y = y_1 + y_{31}t + y_2 s + y_{31}st - y_1 s - y_{42}st$$

But since edges $\overline{P_1P_3}$ and $\overline{P_2P_4}$ are parallel, $x_{31} = x_{42}$ and $y_{31} = y_{42}$. This makes the cross terms cancel out and the remaining equations are:

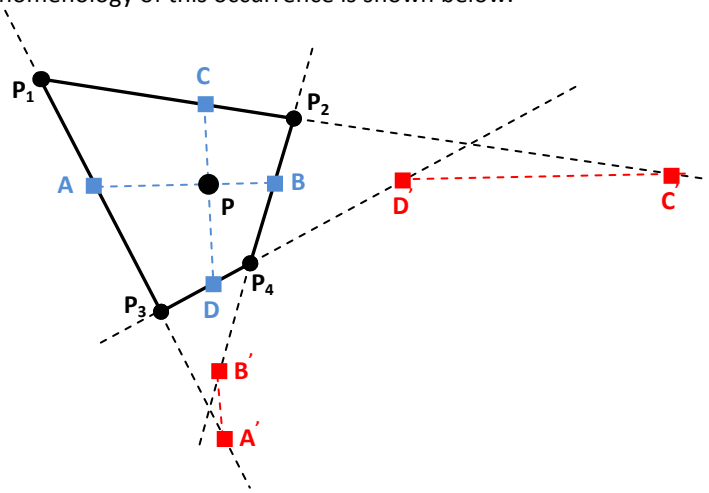
$$P_x = x_1 + x_{31}t + (x_2 - x_1)s$$
$$P_y = y_1 + y_{31}t + (y_2 - y_1)s$$

This system can now be solved as a set of linear equations.

$$\begin{bmatrix} x_{21} & x_{31} \\ y_{21} & y_{31} \end{bmatrix} \cdot \begin{bmatrix} s \\ t \end{bmatrix} = \begin{bmatrix} P_x - x_1 \\ P_y - y_1 \end{bmatrix}$$

Error Conditions

As mentioned earlier, the quadratic solution for the s and t variables presents the problem of solving for the incorrect roots. The phenomenology of this occurrence is shown below.



The blue dashed lines and boxes indicate the results from the correct roots. The point P is solved for using values of s and t that are between 0 and 1. The red boxes and dashed lines show the results from the other solution to this quadratic. At the values of A', B', C', and D' the lines formed still intersect at point P, but the values make less sense. In particular the horizontal intersection line $\overline{AB}$ has now become the vertical and vise versa.

Another problem that can arise is the condition of parallel lines in the interpolation region. This condition is what

drives the need for multiple solution methods. Of course, there is the condition of a rectangular region, or more generally a parallelogram, which would cause both methods to fail. This condition must be checked for explicitly.

For the vertical edges of the search region, the parallelism condition can be checked for using the cross product of the two vertical edges.

$$\text{If} \quad (x_3 - x_1)(y_4 - y_2) - (x_4 - x_2)(y_3 - y_1) == 0 \quad \text{then} \quad VertLines = Parallel$$

Likewise for the horizontal edges the cross product can be used to determine parallelism.

$$\text{If} \quad (x_2 - x_1)(y_4 - y_3) - (x_4 - x_3)(y_2 - y_1) == 0 \quad \text{then} \quad HorzLines = Parallel$$

The processing structure uses this information to determine which method to use when performing the interpolation. The processing logic is as follows:

$$\text{if } (VertLines \neq Parallel) \ \& \ (HorzLines \neq Parallel) \ \rightarrow \text{use Method 1}$$
$$\text{if } (VertLines \neq Parallel) \ \& \ (HorzLines = Parallel) \ \rightarrow \text{use Method 1}$$
$$\text{if } (VertLines = Parallel) \ \& \ (HorzLines \neq Parallel) \ \rightarrow \text{use Method 2}$$
$$\text{if } (VertLines = Parallel) \ \& \ (HorzLines = Parallel) \ \rightarrow \text{use Method 3}$$