as evidenced by Figure 55.5, the exact solution is

$$w = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad b = -1,$$

and for $K = 0.01$, we find that

$$w = \begin{pmatrix} 0.9999 \\ 0.9999 \end{pmatrix}, \quad b = -0.9999.$$

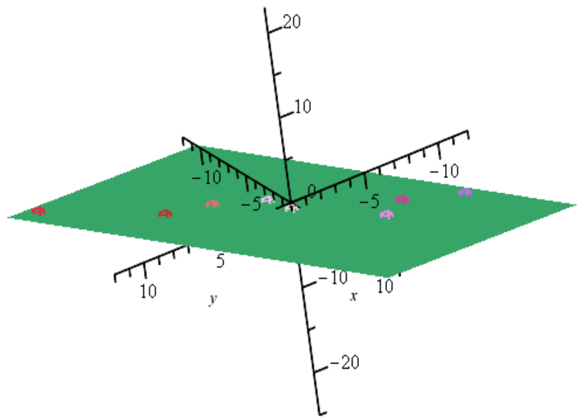

Figure 55.5: The data (X, y_2) of Example 55.1 is contained within the graph of the plane $f(x, y) = x + y - 1$.

We can see how the choice of K affects the quality of the solution (w, b) by computing the norm $\|\xi\|_2$ of the error vector $\xi = \widehat{y} - \widehat{X}w$. We notice that the smaller K is, the smaller is this norm.

It is natural to wonder what happens *if we also penalize* b in program (**RR3**). Let us add the term Kb^2 to the objective function. Then we obtain the program

$$\begin{aligned} &\text{minimize} \quad \xi^\top \xi + K w^\top w + K b^2 \\ &\text{subject to} \\ &\qquad\qquad y - Xw - b\mathbf{1} = \xi, \end{aligned}$$

minimizing over ξ, w and b.

This suggests treating b an an extra component of the weight vector w and by forming the $m \times (n+1)$ matrix $[X\ \mathbf{1}]$ obtained by adding a column of 1's (of dimension m) to the matrix X, we obtain