Our present task is to compute the roots of a function and the gradient of those roots with respect to any auxiliary parameters. Formally, given the function $f : X \times Y \to Z$ and a point $y \in Y$ we want to compute a $x^* \in X$ in the neighborhood of some starting point $x \in X$ satisfying $f(x^*, y) = 0$ and the Jacobian describing how the root changes as y is varied.

The implicit function theorem proves that if f satisfies some weak regularity conditions then in a neighborhood of the starting point (x, y) the roots are given by, well, implicit functions, $x^* = h(y)$, where $h : Y \to X$ satisfies $f(h(y), y) = 0$. In particular, we can then define the desired Jacobian as $\partial h(y)/\partial y$.

Now the implicit function theorem is not directly constructive – it shows only that the functions $h(y)$ exist. To do that, however, it explicit constructs the Jacobian of $g(y)$, which is all we need! Defining the two initial Jacobian matrices

$$J_x(x, y) = \frac{\partial f}{\partial x}(x, y)$$

and

$$J_y = \frac{\partial f}{\partial y}(x, y),$$

then the Jacobian of the roots is given by

$$\frac{\partial h}{\partial y}(y) = - \left[J_x(h(y), y) \right]^{-1} J_y(h(y), y),$$

assuming that J_x is invertible.

The invertibility condition places severe restrictions on the structure of the original constraint. In particular, let $X = R^L$, $Y = R^M$, and $Z = R^N$. Then J_x is a $N \times L$ matrix and is invertible if and only if $N = L$. In other words, we need as many constraints as dimensions of X. Even then, J_x can become singular if the roots are not uniquely defined or other pathological occur.

Assuming that $L = N$, the algorithm will proceed as follows. First, we use a root-finder to compute x^* in the neighborhood of some initial x given a y. Note that I keep saying "neighborhood" as the root-finding may not have a unique solution. Second we compute the full Jacobian $N \times (N + M)$ matrix, $J_{x,y}(x^*, y)$ using autodiff. We then partition the full Jacobian into the $N \times N$ matrix $J_x(x^*, y)$ and the $N \times M$ matrix $J_y(x^*, y)$. Upon inverting the square matrix J_x we can then finally compute the $N \times M$ Jacobian

$$\frac{\partial h}{\partial y}(x^*, y) = \left[J_x(x^*, y) \right]^{-1} \cdot J_y(x^*, y).$$