Using the observation about the cross product, and letting $f(x,y,z) = f(x,y)$ with a similar convention for g, $\nabla f = (y-2x, x, 0)$, $\nabla g = (2x+2y, 2x+2y, 0)$ and so

$$\begin{aligned}
&(y-2x, x, 0) \times (2x+2y, 2x+2y, 0) \\
&= (0, 0, (y-2x)(2x+2y) - x(2x+2y)) = 0
\end{aligned}$$

Thus there are two equations, $x^2 + 2xy + y^2 = 4$ and $4xy - 2y^2 + 6x^2 = 0$. Solving these two yields the points of interest $\left(-\frac{1}{2}, -\frac{3}{2}\right), \left(\frac{1}{2}, \frac{3}{2}\right)$. Both give the same value for f a maximum.

The above generalizes to a general procedure which is described in the following major Theorem. All correct proofs of this theorem will involve some appeal to the implicit function theorem or to fundamental existence theorems from differential equations. A complete proof is very fascinating but it will not come cheap. Good advanced calculus books will usually give a correct proof. If you are interested, there is a complete proof in an appendix to this book. First here is a simple definition explaining one of the terms in the statement of this theorem.

Definition 9.4. Let A be an $m \times n$ matrix. A submatrix is any matrix which can be obtained from A by deleting some rows and some columns.

Theorem 9.4. *Let U be an open subset of $\mathbb{R}^n$ and let $f : U \to \mathbb{R}$ be a C^1 function. Then if $\mathbf{x}_0 \in U$, has the property that*

$$g_i(\mathbf{x}_0) = 0, \ i = 1, \cdots, m, \ g_i \text{ a } C^1 \text{ function,} \tag{9.2}$$

and $\mathbf{x}_0$ is either a local maximum or local minimum of f on the intersection of the level sets just described, and if some $m \times m$ submatrix of

$$D\mathbf{g}(\mathbf{x}_0) \equiv \begin{pmatrix} g_{1x_1}(\mathbf{x}_0) & g_{1x_2}(\mathbf{x}_0) & \cdots & g_{1x_n}(\mathbf{x}_0) \\ \vdots & \vdots & & \vdots \\ g_{mx_1}(\mathbf{x}_0) & g_{mx_2}(\mathbf{x}_0) & \cdots & g_{mx_n}(\mathbf{x}_0) \end{pmatrix}$$

has nonzero determinant, then there exist scalars, $\lambda_1, \cdots, \lambda_m$ such that

$$\begin{pmatrix} f_{x_1}(\mathbf{x}_0) \\ \vdots \\ f_{x_n}(\mathbf{x}_0) \end{pmatrix} = \lambda_1 \begin{pmatrix} g_{1x_1}(\mathbf{x}_0) \\ \vdots \\ g_{1x_n}(\mathbf{x}_0) \end{pmatrix} + \cdots + \lambda_m \begin{pmatrix} g_{mx_1}(\mathbf{x}_0) \\ \vdots \\ g_{mx_n}(\mathbf{x}_0) \end{pmatrix} \tag{9.3}$$

holds.

To help remember how to use (9.3), do the following. First write the Lagrangian,

$$L = f(\mathbf{x}) - \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x})$$

and then proceed to take derivatives with respect to each of the components of $\mathbf{x}$ and also derivatives with respect to each λ_i and set all of these equations equal to 0. The formula (9.3) is what results from taking the derivatives of L with respect to