Solving over- and under-determined sets of equations

Suppose

$$\mathbf{y} = M\mathbf{x}$$

where M is a $n \times m$ matrix, $\mathbf{y}$ is a known n-vector and $\mathbf{x}$ is an unknown m-vector.

First, assume $n > m$. In this case there are more constraints than unknowns, and the system is overdetermined, with no solutions (except for degenerate cases). We can find a least-squares solution that minimizes the error $(\mathbf{y} - M\mathbf{x})$. We want to find $\mathbf{x}$ that minimizes

$$\|\mathbf{y} - M\mathbf{x}\|^2$$

or

$$(\mathbf{y} - M\mathbf{x})^T(\mathbf{y} - M\mathbf{x})$$

or

$$\mathbf{y}^T\mathbf{y} - \mathbf{y}^T M\mathbf{x} - \mathbf{x}^T M^T\mathbf{y} + \mathbf{x}^T M^T M\mathbf{x}$$

Differentiating w.r.t. $\mathbf{x}$ and setting the result equal to zero yields

$$-(\mathbf{y}^T M)^T - (M^T\mathbf{y}) + 2M^T M\mathbf{x} = 0$$

so

$$\mathbf{x} = (M^T M)^{-1} M^T\mathbf{y}$$

where $(M^T M)^{-1} M^T$ (a $m \times n$ matrix) is called a pseudo-inverse.

Second, assume $n < m$. In this case there are fewer constraints than unknowns, and the system is underdetermined, with an infinite number of solutions. We can pick one of these solutions by finding the smallest one. That is, we will minimize $\mathbf{x}$ subject to the constraint $\mathbf{y} = M\mathbf{x}$. The method of Lagrange multipliers has us add a term to the quantity to be minimized:

$$\|\mathbf{x}\|^2 + \boldsymbol{\lambda}^T(\mathbf{y} - M\mathbf{x})$$

Differentiting w.r.t $\mathbf{x}$ and setting the result equal to zero yields

$$2\mathbf{x} - M^T\boldsymbol{\lambda} = 0$$

We can't just solve for $\boldsymbol{\lambda}$ since M is not a square matrix, but we can premultiply by M to obtain

$$2M\mathbf{x} - MM^T\boldsymbol{\lambda} = 0$$

and using $\mathbf{y} = M\mathbf{x}$ gives us

$$2\mathbf{y} = MM^T\boldsymbol{\lambda}$$

so

$$\boldsymbol{\lambda} = 2(MM^T)^{-1}\mathbf{y}$$

and hence

$$\mathbf{x} = M^T(MM^T)^{-1}\mathbf{y}$$

where $M^T(MM^T)$ (a $m \times n$ matrix) is called a pseudo-inverse.