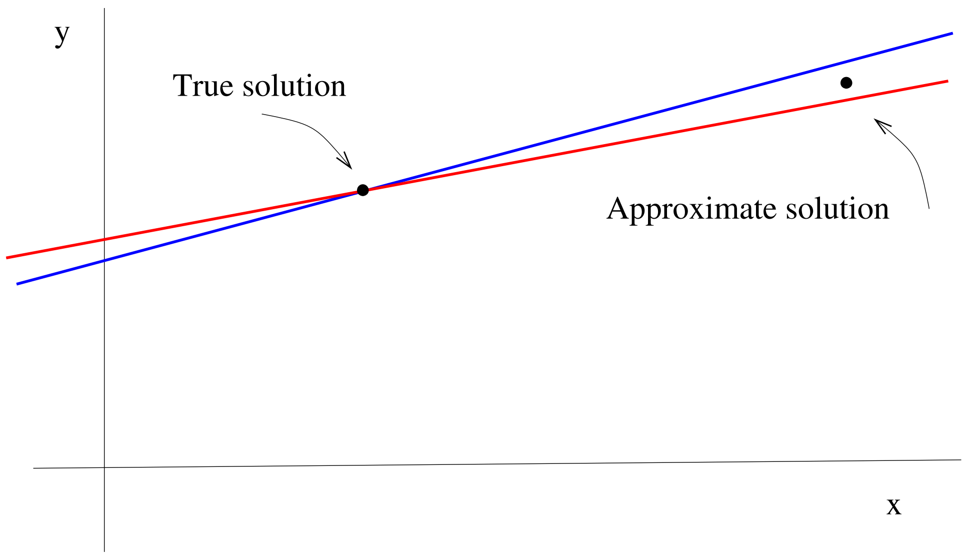
# CSC 721 Algorithms Fall 2017

## Regularization

### Ill-Conditioning

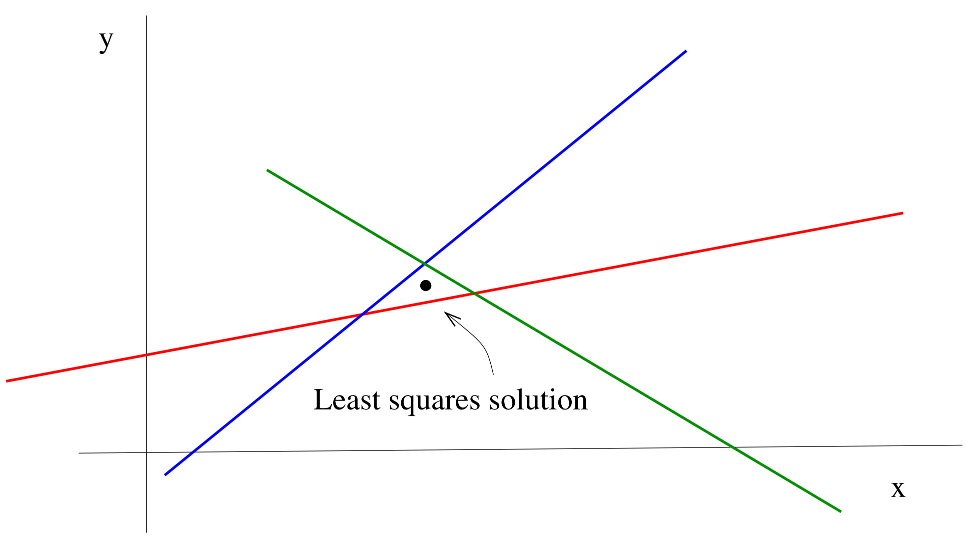



In the diagram above, the approximate solution point is (nearly) on both lines (i.e., a solution). If we substitute the approximate solution into the equations of both lines, it will appear to closely approximate a point that satisfies both equations.

But, appearing to (nearly) satisfy both equations can be misleading. When the lines are nearly parallel, the approximate solution can be far away from the true solution and still appear to (nearly) satisfy the equations. A similar phenomenon occurs in larger systems of equations.

### Over Determined Systems and the Normal Equations

An over-determined system is one in which there are more equations than unknowns. A 2-D example with three equations is illustrated below.



Let $Ax = b$ be a system of equations where $A$ is an $m \times n$ matrix and $m > n$ (i.e., more equations than unknowns). The least squares solution is defined by:

$$\hat{x} = \text{argmin}_x ||Ax - b||_2^2$$

I.e., we minimize the 2-norm of the residual.

Using standard calculus methods, we can derive the so-called **normal equations**. The least squares solution satisfies:

$$A^{\mathrm{T}} A\, x = A^{\mathrm{T}} b$$

Notice that $A^{\mathrm{T}} A$ is an $n \times n$ square matrix and $A^{\mathrm{T}} b$ is an $n \times 1$ vector.

Unfortunately, forming the normal equations makes the ill-conditioning much worse. For some problems, methods such as QR factorization which avoid the normal equations are preferred. However, if you have an ill-conditioned square matrix, regularization is often the best choice.

## Regularization

In general, regularization uses a perturbation of the ill-conditioned problem at hand. If you have a problem you can't solve, you make a "nearby problem" that you can solve.

**Tikhonov Regularization** The following method can be used to improve the conditioning of a linear problem. We form the equations:

$$\begin{bmatrix} A \\ \alpha I \end{bmatrix} x = \begin{bmatrix} b \\ 0 \end{bmatrix} \tag{1}$$

The corresponding optimization problem is:

$$\text{argmin}_x \left\{ \hat{x} = ||Ax - b||_2^2 + ||\alpha x||_2^2 \right\} \tag{2}$$

In equation (2), the term $||\alpha x||_2^2$ acts as a penalty term to constrain $x$ from becoming excessively large (in the sense of the 2-norm).