Lecture 5

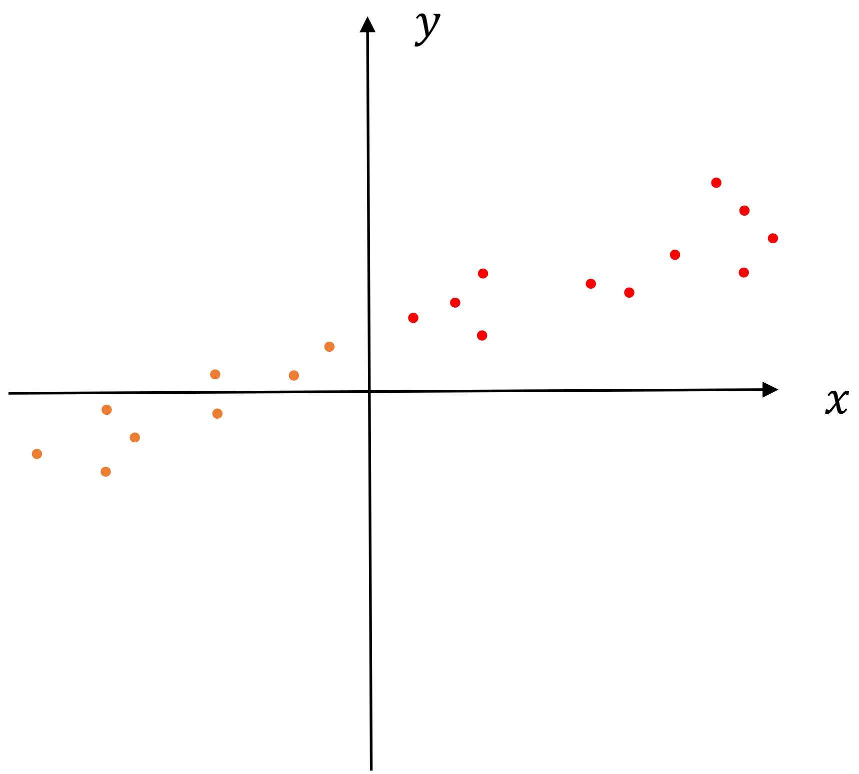
Least squares estimation
(regression vs. total least squares)

Examples: lines & vanishing points

Example Problem 1

How to fit a line to a set of points (x_i, y_i)?

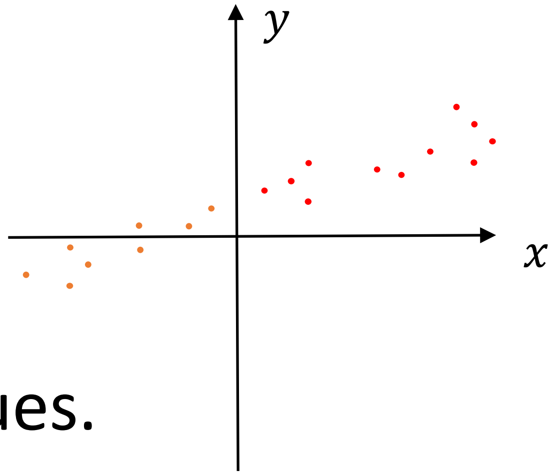
These could be locations of edges (with orientation ignored).



Least squares: version 1 (linear regression)

Model is:

$$y_i = m\, x_i + c + \; n_i$$



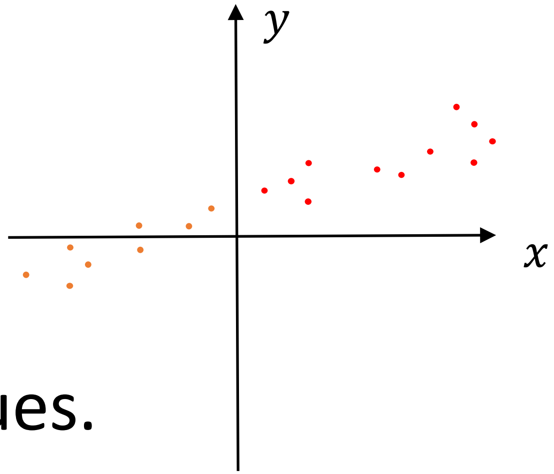
where n_i is additive noise in the y_i values.

We want to solve for m and c. How ?

Least squares: version 1 (linear regression)

Model is:

$$y_i = m\,x_i + c + n_i$$



where n_i is additive noise in the y_i values.

We solve for:

$$\underset{m,\,c}{\operatorname{argmin}} \sum_i (y_i - (m\,x_i + c))^2$$

“Why least *squares*?”

Two reasons:

1. It is mathematically convenient because it allows one to solve minimization problems (e.g. find m, c) using linear algebra.

2. If the noise n_i has a Gaussian probability density, then one can show that minimizing the squared errors gives the best estimate of m, b in a “maximum likelihood” sense.

[ASIDE: I am omitting technical details about maximum likelihood here because it would be distracting for those who don’t know what it is. The term “likelihood” refers to a particular is a conditional probability. The maximum likelihood method maximizes that conditional probability.]

Take derivative of $\sum_i (y_i - (m\, x_i + c))^2$ w.r.t. c and set to 0:

$$\sum_i 2\,(y_i - (m\, x_i + c)) \;=\; 0$$

Take derivative of $\sum_i (y_i - (m\, x_i + c))^2$ w.r.t. m and set to 0:

$$\sum_i 2\,(y_i - (m\, x_i + c))\; x_i \;=\; 0$$

So we have two linear equations with two unknowns and we can solve for m, c.

[This method goes back to Gauss in early 1800's.]

$$\begin{bmatrix} \sum_{i=1}^{N} x_i^2 & \sum_{i=1}^{N} x_i \\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} 1 \end{bmatrix} \begin{bmatrix} m \\ c \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} x_i \, y_i \\ \sum_{i=1}^{N} y_i \end{bmatrix}$$

More generally…… many least squares problems can be written as follows:

Given $\boldsymbol{A}$ and $\boldsymbol{b}$ defined below, find the $\boldsymbol{u}$ that minimizes:

$$\| \boldsymbol{A}\,\boldsymbol{u} - \boldsymbol{b} \|^2$$

where

$\boldsymbol{A}$ is an $m \times n$ matrix

$\boldsymbol{u}$ is a $n \times 1$ vector of variables

$\boldsymbol{b}$ is a $m \times 1$ data vector

In the case of the linear regression problem (fitting a line), the idea is that we stack all the equations

$$y_i = m\, x_i + c$$

and write them as follows:

$$\begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{bmatrix} \begin{bmatrix} m \\ c \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

which is of the form $\boldsymbol{A}\,\boldsymbol{u} = \boldsymbol{b}$ used in the previous slide.

For the more general formulation, we can solve it as follows. We want to minimize the expression (L2 norm) on the left side, so we expand:

$$\| \mathbf{Au} - \mathbf{b} \|^2 \;=\; (\mathbf{Au} - \mathbf{b})^T(\mathbf{Au} - \mathbf{b})$$

$$=\; \mathbf{u}^T\mathbf{A}^T\mathbf{Au} - 2\mathbf{b}^T\mathbf{Au} + \mathbf{b}^T\mathbf{b}$$

Take the derivative with respect to each of the ***u*** variables and set to 0. This gives:

$$2\mathbf{A}^T\mathbf{Au} - 2\mathbf{A}^T\mathbf{b} = 0$$

$$\mathbf{A}^T\mathbf{Au} = \mathbf{A}^T\mathbf{b}$$

If the columns are of A are linearly independent, then we can invert the matrix on the left side to get:

$$\mathbf{u} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$$

So, to find the u that minimizes:

$$\| \mathbf{Au} - \mathbf{b} \|^2 = (\mathbf{Au} - \mathbf{b})^T (\mathbf{Au} - \mathbf{b})$$

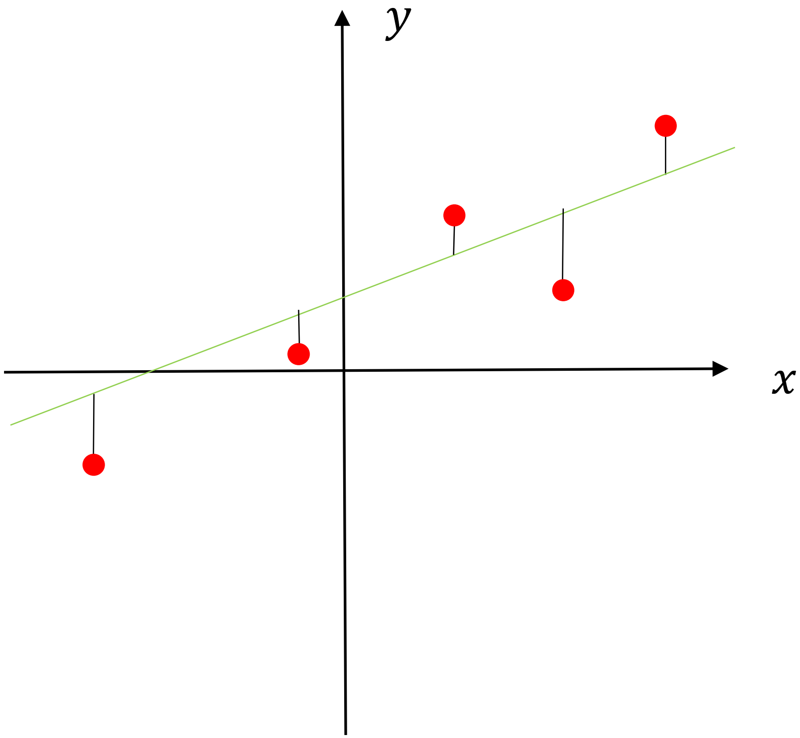
we compute:

$$\mathbf{u} = \underbrace{(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T}\mathbf{b}$$

This matrix is called the *pseudoinverse* of **A**.

Version 1: linear regression

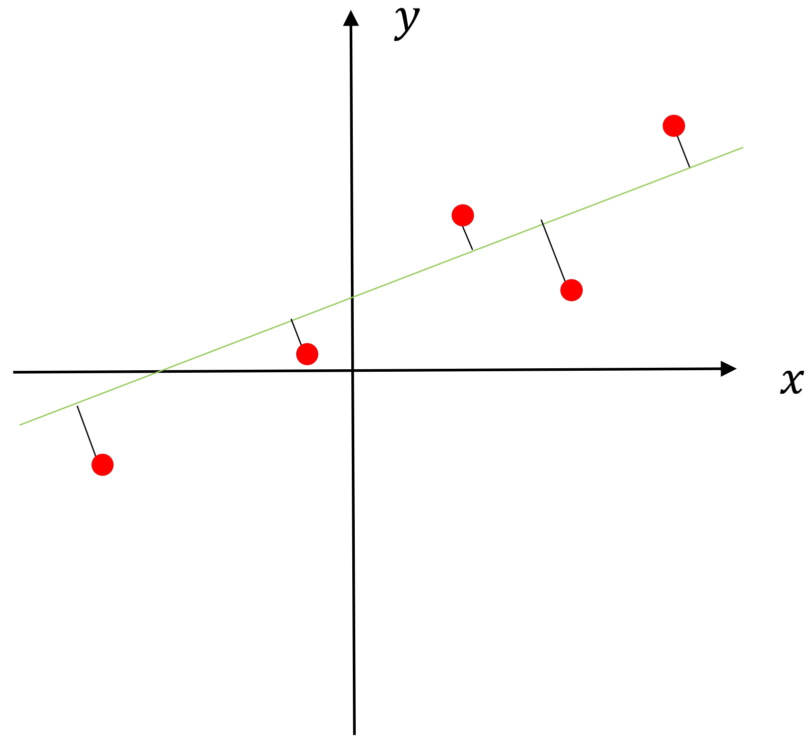
Error is distance to line in y direction only.



$$y = mx + c$$

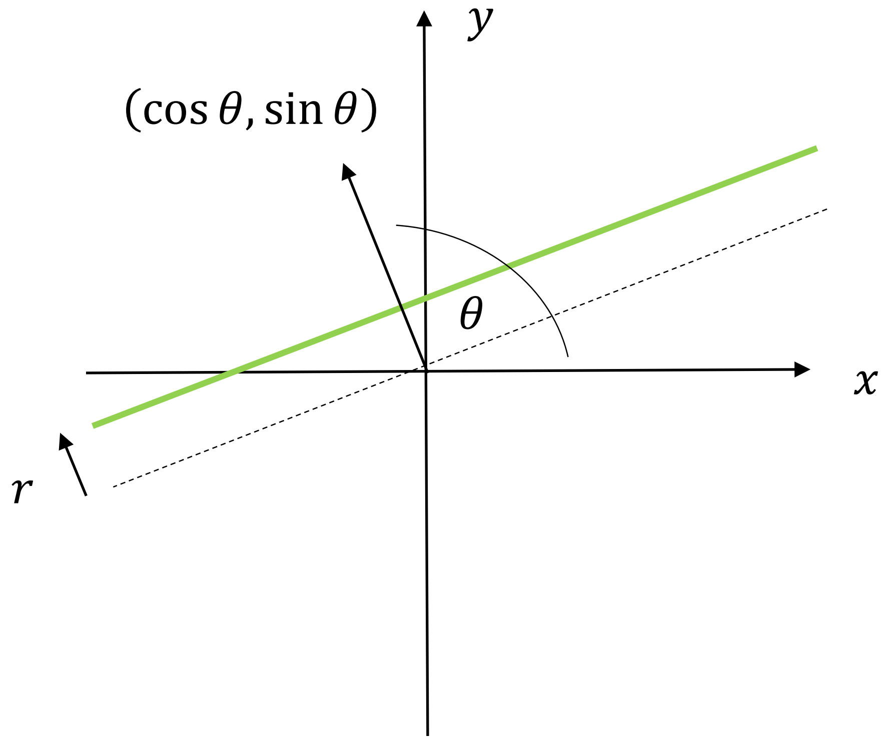
Version 2: "total least squares"

Error is distance perpendicular to line.



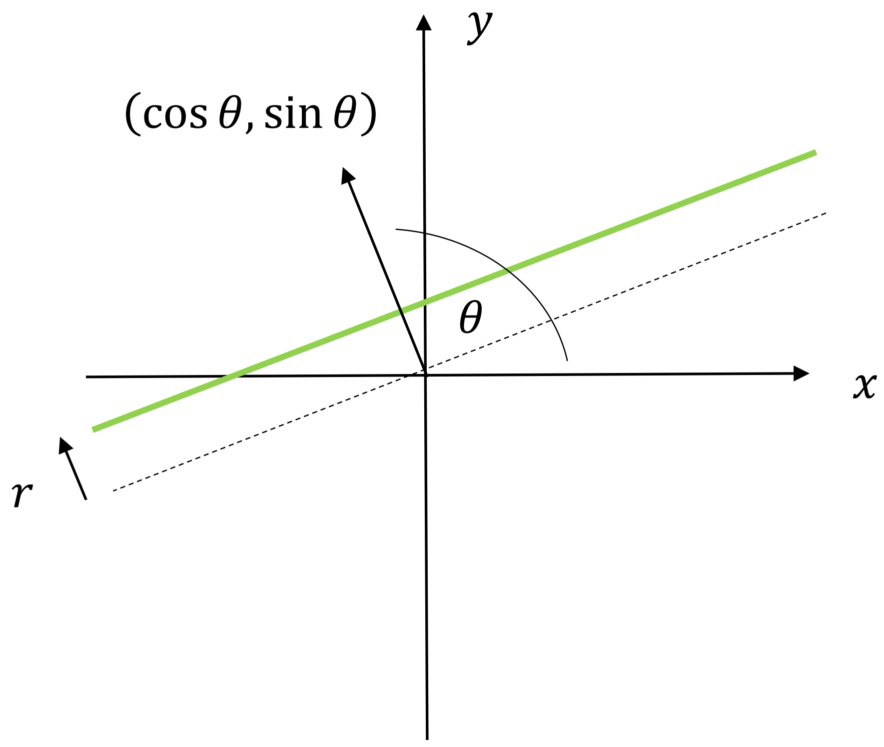
We need a different line representation for perpendicular distances.

A line can be represented using an angle θ in [0, 360) degrees, and a perpendicular distance $r \geq 0$ of the line away from the origin.



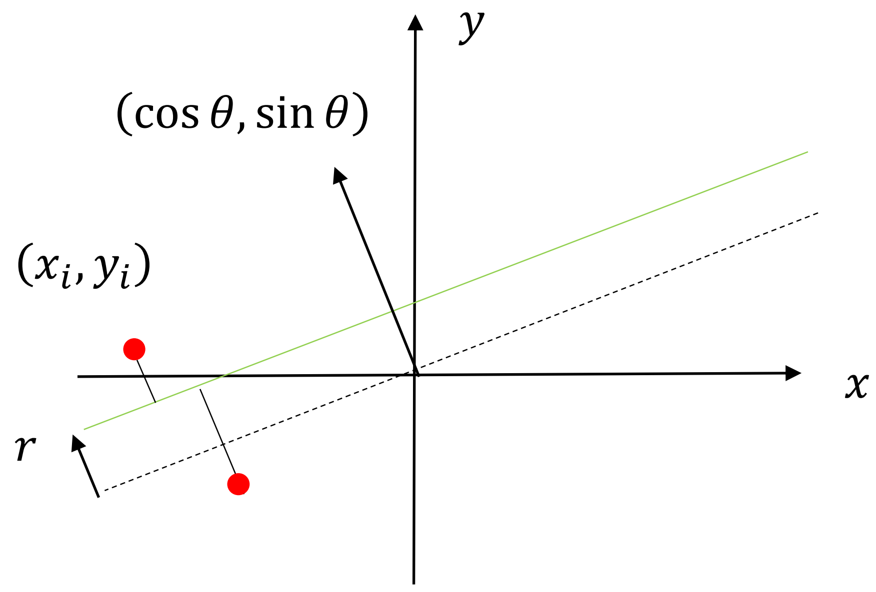
How ?

A line can be represented using an angle θ in [0, 360) degrees, and a perpendicular distance $r \geq 0$ of the line away from the origin.

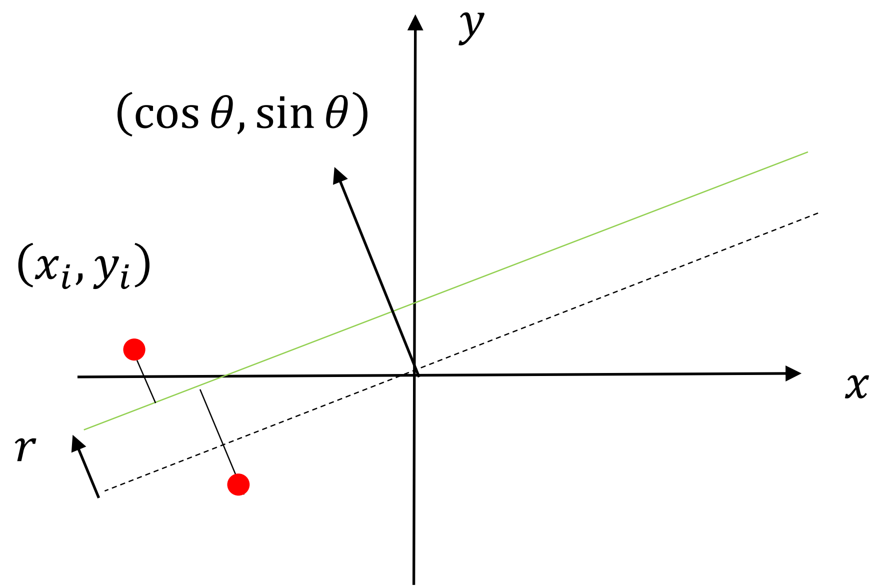


$$x \cos\theta + y \sin\theta = r$$

What is the perpendicular distance from a point (x_i, y_i) to the line given by parameters (θ, r) ?



What is the perpendicular distance from a point (x_i, y_i) to the line given by parameters (θ, r) ?



The distance from (x_i, y_i) to the dashed line through the origin is

$$| \; x_i \cos\theta + y_i \sin\theta \; |.$$

Thus, the distance from (x_i, y_i) to the model line that is a distance r is

$$| \; x_i \cos\theta + y_i \sin\theta \; - \; r \; |$$

To find the best fitting line, we can solve:

$$\underset{\theta, r}{\operatorname{argmin}} \sum_i (x_i \cos\theta + y_i \sin\theta \;-\; r)^2$$

How ?

To find the best fitting line, we can solve:

$$\underset{\theta, r}{\operatorname{argmin}} \sum_i (\, x_i \cos\theta + y_i \sin\theta \;\; - \; r\,)^2$$

First, take derivative with respect to r and set it to 0. This gives:

$$0 = \sum_i (\, x_i \cos\theta + y_i \sin\theta \;\; - \; r)$$

So, $\quad 0 = \bar{x}\cos\theta + \bar{y}\sin\theta \; - r$

where $(\bar{x}, \bar{y}) = \; \frac{1}{N}\sum_{i=1}^{N} \; (x_i, y_i)$.

This mean value $(\bar{x}, \bar{y})$ lies on the best fitting line.

Using the result on the previous slide, i.e. substituting for r, we want to find:

$$\underset{\theta}{\operatorname{argmin}} \sum_i ((x_i - \bar{x}) \cos\theta + (y_i - \bar{y}) \sin\theta)^2$$

How do we find this θ using linear algebra?

[Taking derivative with respect to θ and set to 0 is awkward.]

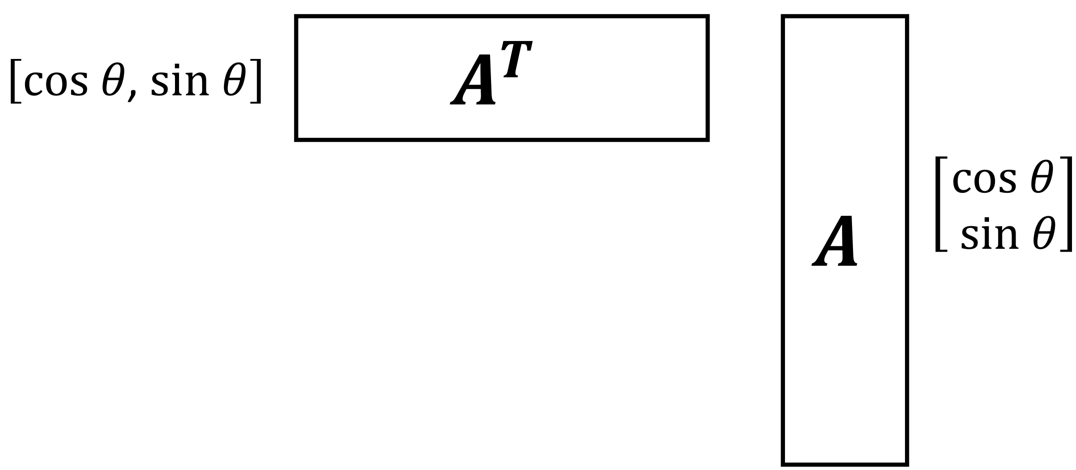
We can write:

$$\sum_i \left((x_i - \bar{x}) \cos\theta + (y_i - \bar{y}) \sin\theta \right)^2$$

as a matrix product:

$$[\cos\theta, \sin\theta] \; \boldsymbol{A}^{\boldsymbol{T}} \boldsymbol{A} \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$$

where $\boldsymbol{A}$ is an $N \; x \; 2$ matrix with rows $(x_i - \bar{x}, \; y_i - \bar{y})$ for $i = 1 \; to \; N$.



$$[\cos\theta, \sin\theta] \;\; \boldsymbol{A^T A} \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \;\; = \;\; \left\| \boldsymbol{A} \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \right\|^2 \;\; \geq \;\; 0$$

We want to find θ that minimizes this quantity. How ?

Compute the eigenvectors and eigenvalues of $\boldsymbol{A^T A}$ (2 x 2 matrix).

Take the unit eigenvector which has the smaller eigenvalue.

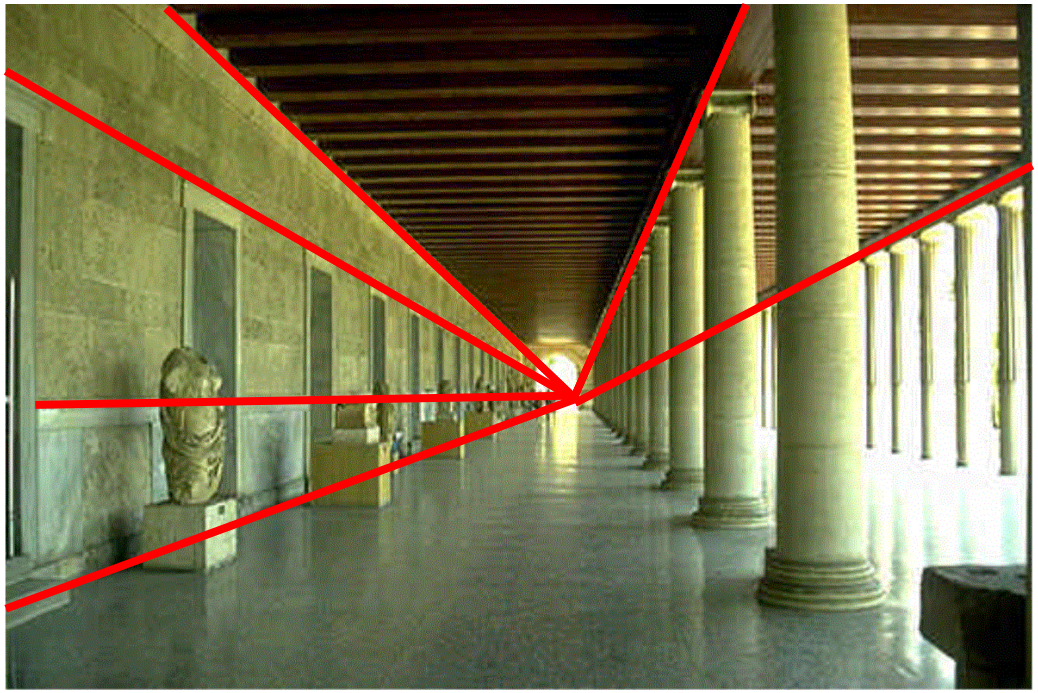
Take the θ such that $[\cos\theta, \sin\theta]^T$ is this unit eigenvector.

Later in the course, I will present a more general formulation.

Vanishing points

Parallel lines in the scenes meet "at infinity". The location of this meeting point in the image is called a vanishing point.
(We will give a more formal presentation of the mathematics of vanishing points in the second half of the course.)

Vanishing points



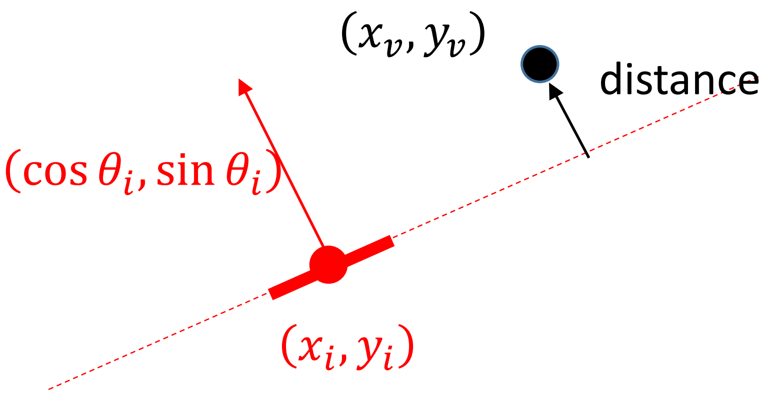
Here is another example.

Vanishing point detection

(output of Canny edge detector)

Suppose we have a set of edge estimates $(x_i, y_i,\ \theta_i)$ where θ_i indicates direction *perpendicular* to the edge, same as last lecture.

We would like to estimate a vanishing point (x_v, y_v) based on these edge estimates. Note: the problem is difficult because most of the edges do NOT point to the vanishing point.



For each edge estimate $(x_i, y_i,\ \theta_i)$ and for any hypothetic vanishing point (x_v, y_v), what is the perpendicular distance from the vanishing point to the line defined by that edge?

The perpendicular distance from the vanishing point (x_v, y_v) to the line defined by the edge $(x_i\,, y_i\,, \theta_i)$ is:

$$| \, (\, x_i - x_v, \;\; y_i - y_v \,) \; \cdot \; (\cos\theta_i \,, \sin\theta_i) \; |$$

To find the vanishing point, we could minimize the sum of squared distances over all edges:

$$\underset{(\,x_v\,, y_v\,)}{argmin} \; \sum_{i=1}^{N} \Big((\, x_i - x_v, \;\; y_i - y_v \,) \; \cdot \; (\cos\theta_i, \sin_{\theta_i}) \Big)^2$$

Verify for yourself that the sum of squares on the previous slide can be written

$$\| \boldsymbol{A}\, \boldsymbol{x} - \boldsymbol{b} \|^2$$

where

$\boldsymbol{A}$ is an N x 2 matrix with row $[\cos\theta_i, \sin\theta_i]$

$\boldsymbol{x}$ is a 2 x 1 vector (x_v, y_v)

$\boldsymbol{b}$ is an N x 1 vector of values $(x_i,\ y_i) \cdot (\cos\theta_i, \sin\theta_i)$.

Vanishing point detection and “noise”

(output of Canny edge detector)

There are often many edges that do not correspond to the vanishing points.
We don’t want to include these “outlier” edges in our least squares estimates. What to do? (next lecture)

Big Challenge in Vanishing point detection (next lecture)