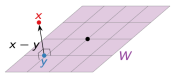
MA1553

Section 6.3

Orthogonal Projections (will finish in next set of slides)

Best Approximation

Suppose you measure a data point x which you know for theoretical reasons must lie on a subspace W.



Due to measurement error, though, the measured x is not actually in W. Best approximation: y is the *closest* point to x on W.

How do you know that y is the closest point? The vector from y to x is orthogonal to W: it is in the *orthogonal complement* $W^\perp$.

Orthogonal Decomposition

Theorem
Every vector x in $\mathbf{R}^n$ can be written as

$$x = x_W + x_{W^\perp}$$

for unique vectors x_W in W and $x_{W^\perp}$ in $W^\perp$.

The equation $x = x_W + x_{W^\perp}$ is called the **orthogonal decomposition** of x (with respect to W).

The vector x_W is the **orthogonal projection** of x onto W.

https://textbooks.math.gatech.edu/ila/demos/projection.html?u1=1,0,0&u2=0,1,0&vec=-1,1,2,1.5&range=3&mode=decomp&closed

Orthogonal Decomposition

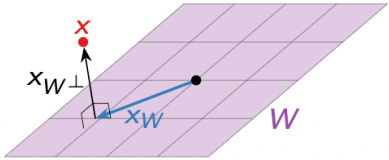
Example

Let W be the xy-plane in $\mathbf{R}^3$. Then $W^\perp$ is the z-axis.

Find the orthogonal decomposition of x=[2;1;3]

$$x = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$$

https://textbooks.math.gatech.edu/ila/demos/projection.html?u1=1,0,0&u2=0,1,1,-.2&vec=-1,1,2,1.5&range=3&mode=decomp&closed



Orthogonal Decomposition
Computation?

Problem: Given x and W, how do you compute the decomposition $x = x_W + x_{W^\perp}$?

Observation: It is enough to compute x_W, because $x_{W^\perp} = x - x_W$.

Theorem (The A^TA Trick)

Let W be a subspace of $\mathbf{R}^n$, let $v_1, v_2, \ldots, v_m$ be a spanning set for W (e.g., a basis), and let

$$A = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_m \\ | & | & & | \end{pmatrix}.$$

Then for any x in $\mathbf{R}^n$, the matrix equation

$$A^TAv = A^Tx \quad \text{(in the unknown vector } v\text{)}$$

is consistent, and $x_W = Av$ for any solution v.

Recipe for Computing $x = x_W + x_{W^\perp}$

- Write W as a column space of a matrix A.
- Find a solution v of $A^TAv = A^Tx$ (by row reducing).
- Then $x_W = Av$ and $x_{W^\perp} = x - x_W$.

The A^TA Trick
Example

Problem: Compute the orthogonal projection of a vector $x = (x_1, x_2, x_3)$ in $\mathbf{R}^3$ onto the xy-plane.

The A^TA Trick
Another Example

Problem: Let

$$x = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \qquad W = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \text{ in } \mathbf{R}^3 \mid x_1 - x_2 + x_3 = 0 \right\}.$$

(1) compute the orthogonal decomposition of x for the given subspace W

(2) find the distance from x to the subspace W

Recipe for Computing $x = x_W + x_{W^\perp}$

- Write W as a column space of a matrix A.
- Find a solution v of $A^TAv = A^Tx$ (by row reducing).
- Then $x_W = Av$ and $x_{W^\perp} = x - x_W$.

Recipe for Computing $x = x_W + x_{W^\perp}$

- Write W as a column space of a matrix A.
- Find a solution v of $A^TAv = A^Tx$ (by row reducing).
- Then $x_W = Av$ and $x_{W^\perp} = x - x_W$.

Orthogonal Projection onto a Line

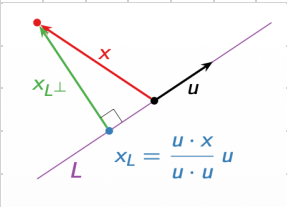
Example

Problem: Compute the orthogonal projection of $x = \begin{pmatrix} -6 \\ 4 \end{pmatrix}$ onto the line L spanned by $u = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$, and find the distance from u to L.

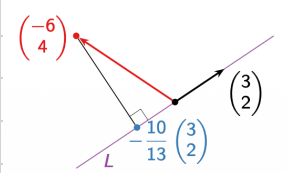
Projection onto a Line

The projection of x onto a line $L = \text{Span}\{u\}$ is

$$x_L = \frac{u \cdot x}{u \cdot u} u \qquad x_{L^\perp} = x - x_L.$$



https://textbooks.math.gatech.edu/ila/demos/projection.html?u1=3,2&vec=-6,4&labels=u&closed&mode=distance



Projection as a Transformation

Change in Perspective: let us consider orthogonal projection as a *transformation*.

Definition
Let W be a subspace of $\mathbf{R}^n$. Define a transformation

$$T\colon \mathbf{R}^n \longrightarrow \mathbf{R}^n \quad \text{by} \quad T(x) = x_W.$$

This transformation is also called **orthogonal projection** with respect to W.

Theorem
Let W be a subspace of $\mathbf{R}^n$ and let $T\colon \mathbf{R}^n \to \mathbf{R}^n$ be the orthogonal projection with respect to W. Then:

1. T is a *linear* transformation.
2. For every x in $\mathbf{R}^n$, $T(x)$ is the *closest* vector to x in W.
3. For every x in W, we have $T(x) = x$.
4. For every x in $W^\perp$, we have $T(x) = 0$.
5. $T \circ T = T$.
6. The range of T is W and the null space of T is $W^\perp$.

Projection Matrix

Example

Problem: Let $L = \text{Span}\left\{\begin{pmatrix}3\\2\end{pmatrix}\right\}$ and let $T\colon \mathbf{R}^2 \to \mathbf{R}^2$ be the orthogonal projection onto L. Compute the matrix A for T.

You Try It!

Poll

Let A be the matrix for proj_W. What is/are the eigenvalue(s) of A?

A. 0 B. 1 C. -1 D. $0, 1$ E. $1, -1$ F. $0, -1$ G. $-1, 0, 1$

Fact:

Theorem
Let W be an m-dimensional subspace of $\mathbf{R}^n$, let $T\colon \mathbf{R}^n \to W$ be the projection, and let A be the matrix for T. Then:

1. $\text{Col}\, A = W$, which is the 1-eigenspace.
2. $\text{Nul}\, A = W^\perp$, which is the 0-eigenspace.
3. $A^2 = A$.
4. A is similar to the diagonal matrix with m ones and $n - m$ zeros on the diagonal.

Projection Matrix
Another Example

Problem: Let

$$W = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \text{ in } \mathbf{R}^3 \mid x_1 - x_2 + x_3 = 0 \right\}$$

and let $T\colon \mathbf{R}^3 \to \mathbf{R}^3$ be orthogonal projection onto W. Compute the matrix B for T.

Solution:

In the slides for the last lecture we computed $W = \operatorname{Col} A$ for

$$A = \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

To compute $T(e_i)$ we have to solve the matrix equation $A^TAv = A^Te_i$. We have

$$A^TA = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \qquad A^Te_i = \text{the } i\text{th column of } A^T = \begin{pmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}.$$

$$\left(\begin{array}{cc|c} 2 & -1 & 1 \\ -1 & 2 & -1 \end{array}\right) \xrightarrow{\text{RREF}} \left(\begin{array}{cc|c} 1 & 0 & 1/3 \\ 0 & 1 & -1/3 \end{array}\right) \implies T(e_1) = \frac{1}{3}A\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}$$

$$\left(\begin{array}{cc|c} 2 & -1 & 1 \\ -1 & 2 & 0 \end{pmatrix} \xrightarrow{\text{RREF}} \left(\begin{array}{cc|c} 1 & 0 & 2/3 \\ 0 & 1 & 1/3 \end{array}\right) \implies T(e_2) = \frac{1}{3}A\begin{pmatrix} 2 \\ 1 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$$

$$\left(\begin{array}{cc|c} 2 & -1 & 0 \\ -1 & 2 & 1 \end{array}\right) \xrightarrow{\text{RREF}} \left(\begin{array}{cc|c} 1 & 0 & 1/3 \\ 0 & 1 & 2/3 \end{array}\right) \implies T(e_2) = \frac{1}{3}A\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} -1 \\ 1 \\ 2 \end{pmatrix}$$

$$\implies B = \frac{1}{3}\begin{pmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 2 \end{pmatrix}.$$

Fact:

Theorem
Let $\{v_1, v_2, \ldots, v_m\}$ be a *linearly independent* set in $\mathbf{R}^n$, and let

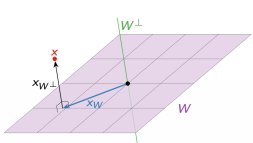
$$A = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_m \\ | & | & & | \end{pmatrix}.$$

Then the $m \times m$ matrix A^TA is invertible.

Summary

Recall: Let W be a subspace of $\mathbf{R}^n$.

- The **orthogonal complement** $W^\perp$ is the set of vectors orthogonal to everything in W.
- The **orthogonal decomposition** of a vector x with respect to W is the unique way of writing $x = x_W + x_{W^\perp}$ for x_W in W and $x_{W^\perp}$ in $W^\perp$.
- The vector x_W is the **orthogonal projection** of x onto W. It is the closest vector to x in W.
- To compute x_W, write W as $\operatorname{Col} A$ and solve $A^TAv = A^Tx$; then $x_W = Av$.



- If $W = L = \operatorname{Span}\{u\}$ is a line then $x_L = \frac{u \cdot x}{u \cdot u}u$.
- Orthogonal projection is a linear transformation.
- The matrix for projection onto W has eigenvalues 1 and 0 with eigenspaces W and $W^\perp$.
- A projection matrix is diagonalizable.

Section 6.5

The Method of Least Squares

Least Squares Solutions

We now are in a position to solve the motivating problem of this third part of the course:

Problem

Suppose that $Ax = b$ does not have a solution. What is the best possible approximate solution?

Let A be an $m \times n$ matrix.

Definition

A **least squares solution** of $Ax = b$ is a vector $\widehat{x}$ in $\mathbf{R}^n$ such that

$$\|b - A\widehat{x}\| \leq \|b - Ax\|$$

Theorem

The least squares solutions of $Ax = b$ are the solutions of

$$(A^T A)\widehat{x} = A^T b.$$

Note we compute $\widehat{x}$ directly, without computing $\widehat{b}$ first.

Least Squares Solutions

Example

Find the least squares solutions of $Ax = b$ where:

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix} \qquad b = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}.$$

Least Squares Solutions

Example, continued

How close did we get?

$$\widehat{b} = A\widehat{x} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} -3 \\ 5 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \\ -1 \end{pmatrix}$$

Main idea:

To say $Ax = b$ does not have a solution means that b is not in Col A.

The closest possible $\widehat{b}$ for which $Ax = \widehat{b}$ does have a solution is $\widehat{b} = b_{\text{Col } A}$.

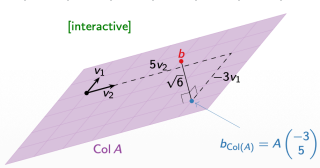
Then $A\widehat{x} = \widehat{b}$ is a consistent equation.

A solution $\widehat{x}$ to $A\widehat{x} = \widehat{b}$ is a **least squares solution.**

How to compute?

Conclusion: $\widehat{x}$ is just a solution of $A^T Av = A^T b$

[interactive]



Note that

$$\begin{pmatrix} -3 \\ 5 \end{pmatrix}$$

records the coefficients of v_1 and v_2 in $\widehat{b}$.

Least Squares Solutions
Second example

Find the least squares solutions of $Ax = b$ where:

$$A = \begin{pmatrix} 2 & 0 \\ -1 & 1 \\ 0 & 2 \end{pmatrix} \qquad b = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

Least Squares Solutions
Uniqueness

When does $Ax = b$ have a *unique* least squares solution $\widehat{x}$?

Theorem
Let A be an $m \times n$ matrix. The following are equivalent:

1. $Ax = b$ has a *unique* least squares solution for all b in $\mathbf{R}^m$.
2. The columns of A are linearly independent.
3. A^TA is invertible.

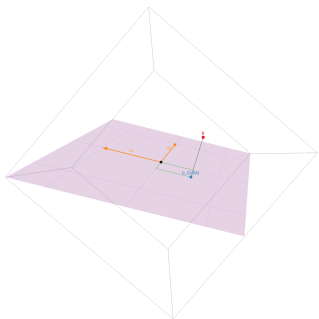
In this case, the least squares solution is $(A^TA)^{-1}(A^Tb)$.

Why? If the columns of A are linearly *d*ependent, then $A\widehat{x} = \widehat{b}$ has many solutions:

Note: A^TA is always a square matrix, but it need not be invertible.

https://textbooks.math.gatech.edu/ila/demos/leastsquares.html?v1=2,-1,0&v2=0,1,2&vec=1,0,-1&range=3





Application
Data modeling: best fit line

Find the best fit line through $(0,6)$, $(1,0)$, and $(2,0)$.

The general equation of a line is

$$y = C + Dx.$$

Key insight:

So we want to solve:

$$\begin{aligned} 6 &= C + D \cdot 0 \\ 0 &= C + D \cdot 1 \\ 0 &= C + D \cdot 2. \end{aligned}$$

Note to self: this is the first example from the lecture.

Summary

- A **least squares solution** of $Ax = b$ is a vector $\widehat{x}$ such that $\widehat{b} = A\widehat{x}$ is *as close to b as possible*.
- This means that $\widehat{b} = b_{\text{Col } A}$.
- One way to compute a least squares solution is by solving the system of equations
$$(A^T A)\widehat{x} = A^T b.$$
Note that $A^T A$ is a (symmetric) square matrix.
- Least-squares solutions are unique when the columns of A are linearly independent.
- You can use least-squares to find best-fit lines, parabolas, ellipses, planes, etc.

Story time ...

Application

Best fit parabola

What least squares problem $Ax = b$ finds the best parabola through the points $(-1, 0.5)$, $(1, -1)$, $(2, -0.5)$, $(3, 2)$?

The general equation for a parabola is

$$y = Ax^2 + Bx + C.$$

So we want to solve:

$$\begin{aligned} 0.5 &= A(-1)^2 + B(-1) + C \\ -1 &= A(1)^2 + B(1) + C \\ -0.5 &= A(2)^2 + B(2) + C \\ 2 &= A(3)^2 + B(3) + C \end{aligned}$$

In matrix form:

$$\begin{pmatrix} 1 & -1 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \end{pmatrix} \begin{pmatrix} A \\ B \\ C \end{pmatrix} = \begin{pmatrix} 0.5 \\ -1 \\ -0.5 \\ 2 \end{pmatrix}.$$

$$M^T M = \begin{pmatrix} 1 & 1 & 4 & 9 \\ -1 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \end{pmatrix} = \begin{pmatrix} 99 & 35 & 15 \\ 35 & 15 & 5 \\ 15 & 5 & 4 \end{pmatrix}$$

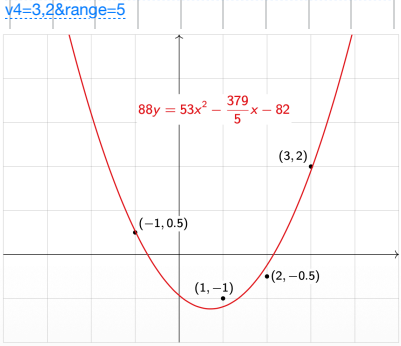
$$M^T b = \begin{pmatrix} 1 & 1 & 4 & 9 \\ -1 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0.5 \\ -1 \\ -0.5 \\ 2 \end{pmatrix} = \begin{pmatrix} \frac{31}{2} \\ \frac{7}{2} \\ 1 \end{pmatrix}$$

$$M^T M = \begin{pmatrix} 99 & 35 & 15 \\ 35 & 15 & 5 \\ 15 & 5 & 4 \end{pmatrix} \quad M^T b = \begin{pmatrix} \frac{31}{2} \\ \frac{7}{2} \\ 1 \end{pmatrix}$$

$$\left[\begin{array}{ccc|c} 99 & 35 & 15 & \frac{31}{2} \\ 35 & 15 & 5 & \frac{7}{2} \\ 15 & 5 & 4 & 1 \end{array}\right] \qquad \left[\begin{array}{ccc|c} 1 & 0 & 0 & \frac{341}{6292} \\ 0 & 1 & 0 & -\frac{4479}{5720} \\ 0 & 0 & 1 & -\frac{41}{44} \end{array}\right]$$

$$A = \frac{341}{6292} \approx 0.0542, \qquad B = -\frac{4479}{5720} \approx -0.78299, \qquad C = -\frac{41}{44} \approx -0.93182.$$

https://textbooks.math.gatech.edu/ila/demos/bestfit.html?func=A*x%5E2+B*x+C&v1=-1,.5&v2=1,-1&v3=2,-.5&v4=3,2&range=5



Application

Best-fit Trigonometric Function

For fun: what is the best-fit function of the form

$$y = A + B\cos(x) + C\sin(x) + D\cos(2x) + E\sin(2x) + F\cos(3x) + G\sin(3x)$$

passing through the points

$$\binom{-4}{-1}, \binom{-3}{0}, \binom{-2}{-1.5}, \binom{-1}{.5}, \binom{0}{1}, \binom{1}{-1}, \binom{2}{-.5}, \binom{3}{2}, \binom{4}{-1}?$$

$$M = \begin{pmatrix} 1 & \cos(-4) & \sin(-4) & \cos(-8) & \sin(-8) & \cos(-12) & \sin(-12) \\ 1 & \cos(-3) & \sin(-3) & \cos(-6) & \sin(-6) & \cos(-9) & \sin(-9) \\ 1 & \cos(-2) & \sin(-2) & \cos(-4) & \sin(-4) & \cos(-6) & \sin(-6) \\ 1 & \cos(-1) & \sin(-1) & \cos(-2) & \sin(-2) & \cos(-3) & \sin(-3) \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & \cos(1) & \sin(1) & \cos(2) & \sin(2) & \cos(3) & \sin(3) \\ 1 & \cos(2) & \sin(2) & \cos(4) & \sin(4) & \cos(6) & \sin(6) \\ 1 & \cos(3) & \sin(3) & \cos(6) & \sin(6) & \cos(9) & \sin(9) \\ 1 & \cos(4) & \sin(4) & \cos(8) & \sin(8) & \cos(12) & \sin(12) \end{pmatrix}, \quad b = \begin{pmatrix} -1 \\ 0 \\ -1.5 \\ 0.5 \\ 1 \\ -1 \\ -0.5 \\ 2 \\ -1 \end{pmatrix}$$

$$M \approx \begin{pmatrix} 1 & -0.65 & 0.76 & -0.15 & -0.99 & 0.84 & 0.54 \\ 1 & -0.99 & 0.14 & 0.96 & 0.28 & -0.91 & -0.41 \\ 1 & -0.42 & -0.91 & -0.65 & 0.76 & 0.96 & 0.28 \\ 1 & 0.54 & -0.84 & -0.42 & -0.91 & -0.99 & -0.14 \\ 1 & 1.00 & 0.00 & 1.00 & 0.00 & 1.00 & 0.00 \\ 1 & 0.54 & 0.84 & -0.42 & 0.91 & -0.99 & 0.14 \\ 1 & -0.42 & 0.91 & -0.65 & -0.76 & 0.96 & -0.28 \\ 1 & -0.99 & -0.14 & 0.96 & -0.28 & -0.91 & 0.41 \\ 1 & -0.65 & -0.76 & -0.15 & 0.99 & 0.84 & -0.54 \end{pmatrix},$$

$$M^T M \approx \begin{pmatrix} 9.00 & -2.04 & 0.00 & 0.33 & 0.00 & -0.20 & 0.00 \\ -2.04 & 4.50 & 0.00 & -1.12 & 0.00 & 0.47 & 0.00 \\ 0.00 & 0.00 & 4.50 & 0.00 & -0.16 & 0.00 & 0.09 \\ 0.33 & -1.12 & 0.00 & 4.50 & 0.00 & -0.55 & 0.00 \\ 0.00 & 0.00 & -0.16 & 0.00 & 4.50 & 0.00 & -0.31 \\ -0.20 & 0.47 & 0.00 & -0.55 & 0.00 & 4.50 & 0.00 \\ 0.00 & 0.00 & 0.09 & 0.00 & -0.31 & 0.00 & 4.50 \end{pmatrix}, \quad M^T b \approx \begin{pmatrix} -1.50 \\ 1.69 \\ -1.55 \\ 0.90 \\ 0.76 \\ -2.29 \\ 2.02 \end{pmatrix}$$

$$\left[\begin{array}{ccccccc|c} 9.00 & -2.04 & 0 & 0.33 & 0 & -0.20 & 0 & -1.50 \\ -2.04 & 4.50 & 0 & -1.12 & 0 & 0.47 & 0 & 1.69 \\ 0 & 0 & 4.50 & 0 & -0.16 & 0 & 0.09 & -1.55 \\ 0.33 & -1.12 & 0 & 4.50 & 0 & -0.55 & 0 & 0.90 \\ 0 & 0 & -0.16 & 0 & 4.50 & 0 & -0.31 & 0.76 \\ -0.20 & 0.47 & 0 & -0.55 & 0 & 4.50 & 0 & -2.29 \\ 0 & 0 & 0.09 & 0 & -0.31 & 0 & 4.50 & 2.02 \end{array}\right] \quad \left[\begin{array}{ccccccc|c} 1 & 0 & 0 & 0 & 0 & 0 & 0 & -0.07 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0.37 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & -0.34 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0.29 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0.16 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -0.47 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0.46 \end{array}\right]$$

$$\begin{aligned} A &\approx -0.07, \\ B &\approx 0.37, \\ C &\approx -0.34, \\ D &\approx 0.29, \\ E &\approx 0.16, \\ F &\approx -0.47, \\ G &\approx 0.46. \end{aligned}$$

https://textbooks.math.gatech.edu/ila/demos/bestfit.html?func=A+B*cos(x)+C*sin(x)+D*cos(2*x)+EE*sin(2*x)+F*cos(3*x)+G*sin(3*x)&v1=-4,-1&v2=-3,0&v3=-2,-1.5&v4=-1,.5&v5=0,1&v6=1,-1&v7=2,-.5&v8=3,2&v9=4,-1&range=5

