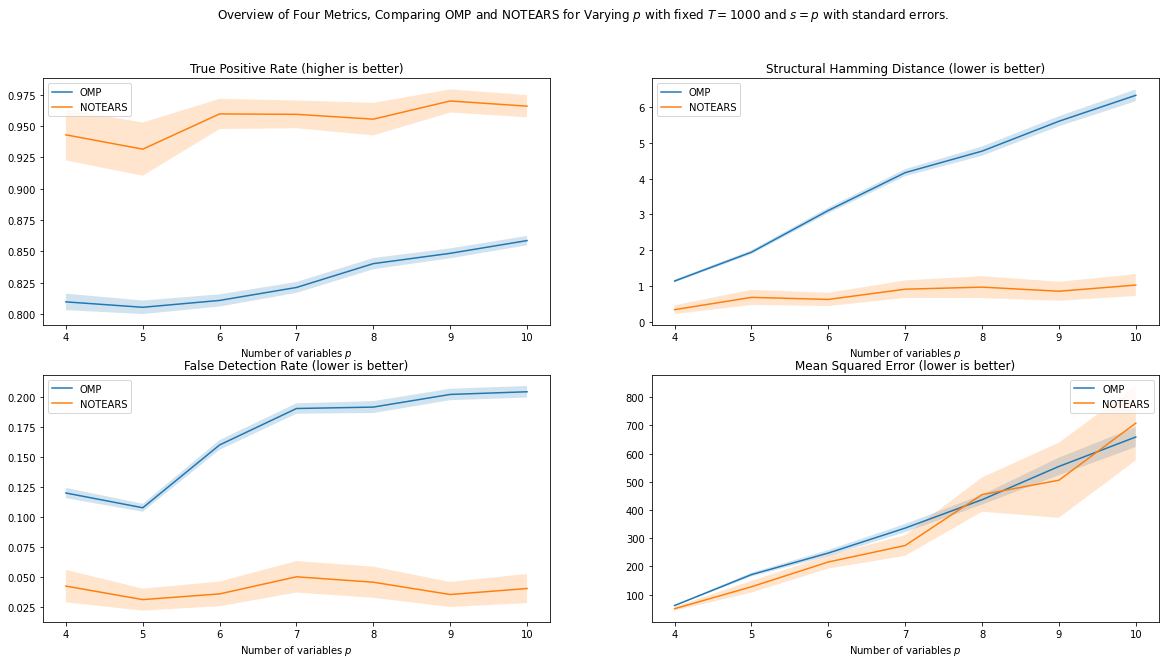
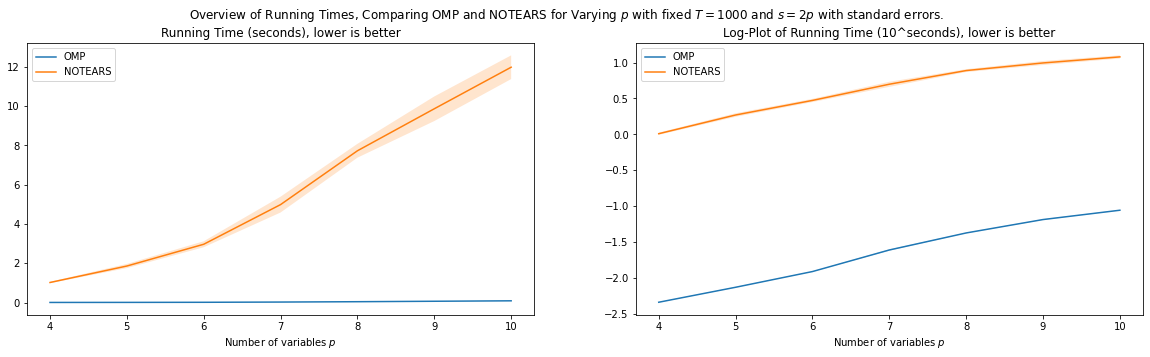
Prep Meeting 21

# SEM for larger dimensions

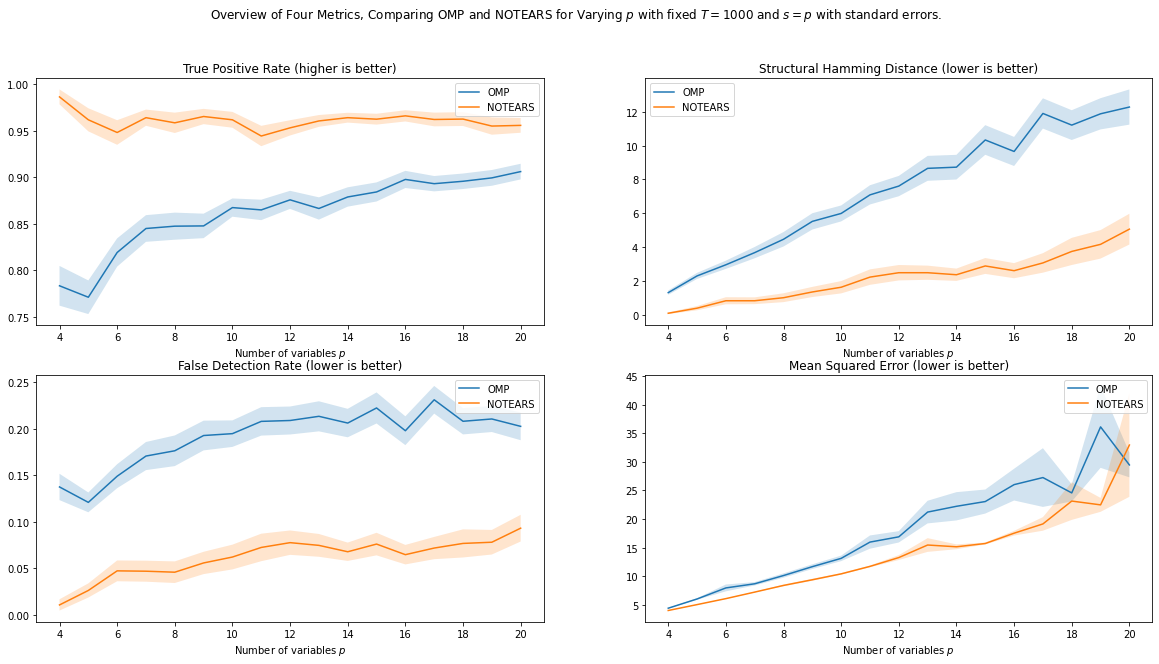
Redid Example for strange *p = 8*. Turns out that it was just an anomaly; some randomly generated SEMs are much harder than others.

Interesting: We see that OMP is not as good as NOTEARS for these small dimensions. Especially the structural hamming distance takes off, mostly accredited to the quite high false detection rate. Furthermore, SHD is an absolute metric, it counts the number of incorrect predictions rather than the rate, so we can expect that to increase, as the number of possible options increases quadratically.

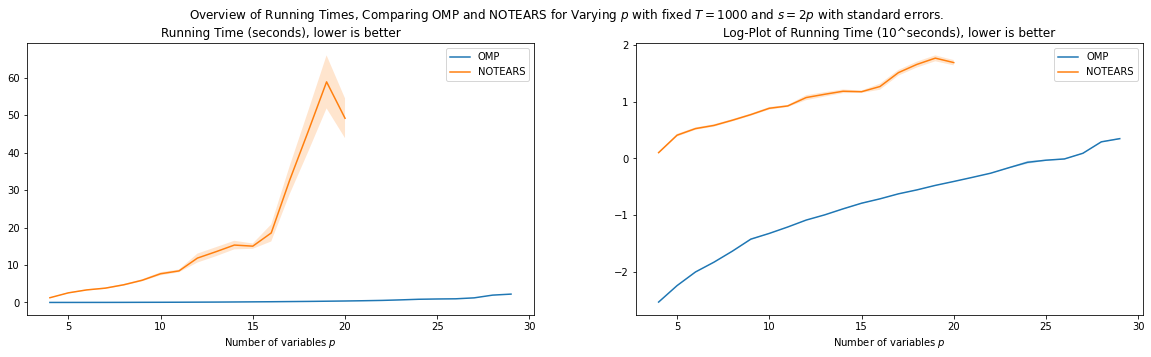


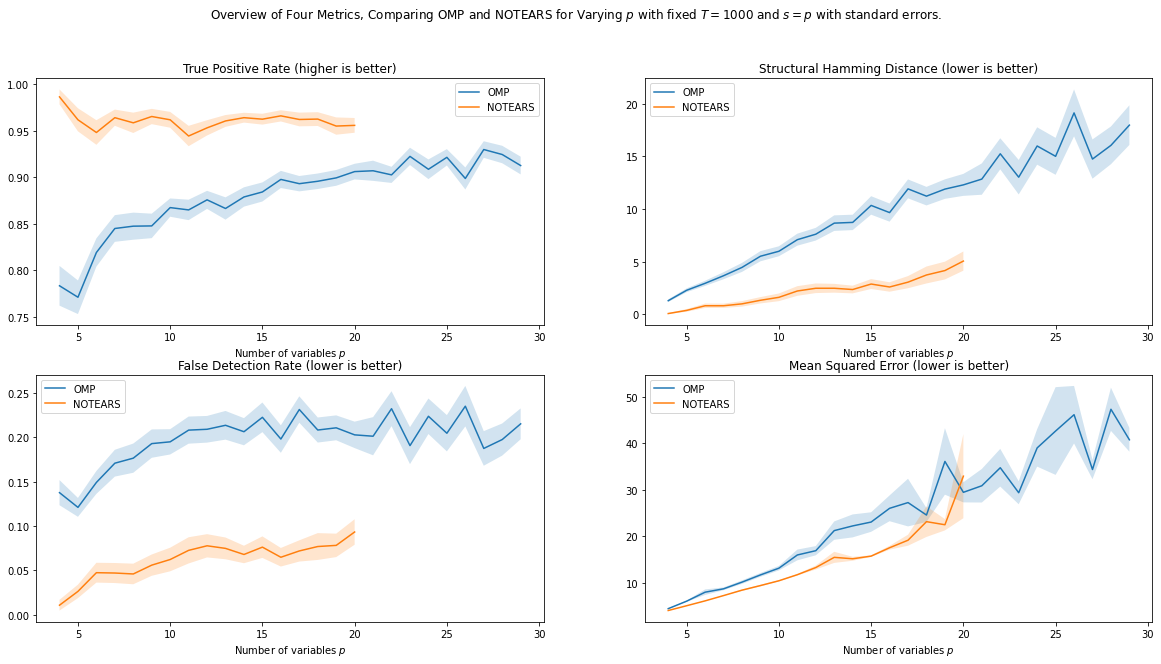


More interesting: Ran NOTEARS for a night, until 20 dimensions. It seems that OMP and other greedy methods do not suffer that much from problems anymore. It seems to get closer to NOTEARS again.



Good question: What happens further, e.g. from 21 – 50? Running would take a couple of nights. NOTEARS takes 1 minute per iteration at 20, so ~ an hour for fifty measurements. Takes multiple days to do the larger analysis.



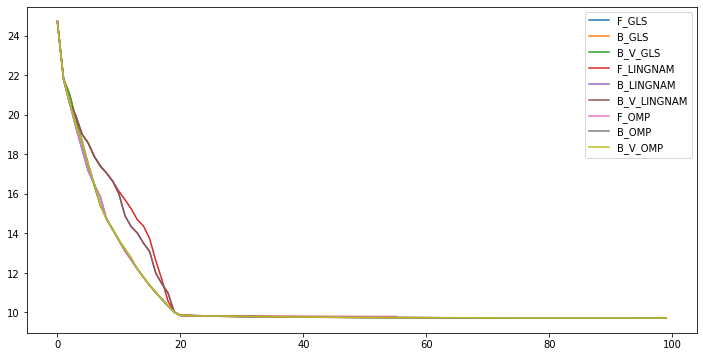


# The Nine Greedy Methods

Implemented all nine greedy methods based on direction and criterion, so

(Forward / Backward / Backward-Violations first) x (OMP / GLS / LINGNAM).

**Results**

****

One example, with *p = 10*. Y-axis is mean squared error on training data. X-axis = #edges, ordered by importance. True edges = 20. *T = 1000*.

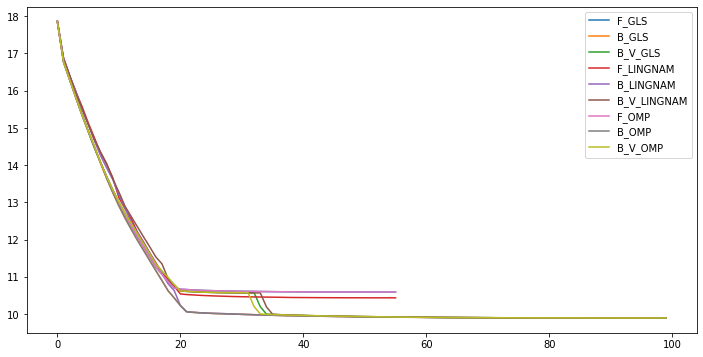
*Methods*

In a simple VAR setting with unit variance, all three methods seem very comparable. LINGNAM is a little bit off sometimes due to its naïve way of adding / removing edges.

OMP and GLS both seem very comparable, which advocates for OMP, as it is quite a bit faster, I recall order of *p* times faster.

*Direction*

The backwards method works well in the scenario of no model mismatch whatsoever. However, when there is even a slight model mismatch, this direction may fail, as we have seen.



The forward and backwards-violations first directions seem to work well, and both often recover all the true edges. They differ slightly in their unregularized form: the forward direction yields a dense DAG, the backwards-violations first direction does often not yield this.

# OMP vs GLS

Interesting question: *In what way do OMP and GLS differ in atom selection?*

**OMP**

OMP picks the atom that is *most correlated* with the current residual.

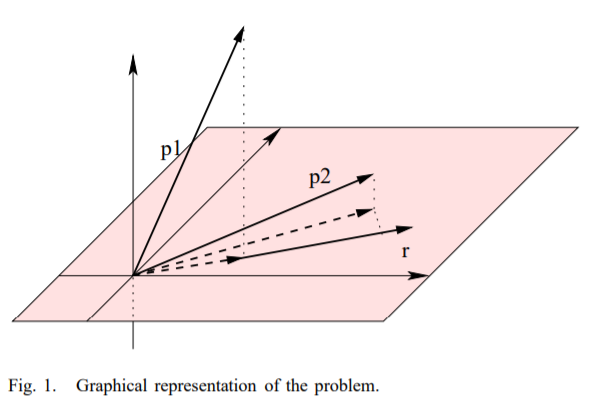
From a geometric perspective, it picks the atom that has the *smallest angle* with the current residual.

**GLS**

GLS picks the atom that yields the *smallest* MSE.

From a geometric perspective, it picks the atom such that its *orthogonal projection with the current residual* yields the *smallest angle*.

**Paper example:**



**OMP** selects the new element based on the inner product, i.e. based on the angle between the vectors and the current residual1 . In the figure, this would be element **p2**.

OLS, on the contrary, selects the element that is able to best approximate the current residual, i.e. OLS will select the element with the *smallest angle after this element is projected into the orthogonal subspace*. The projection of p1 and p2 into the orthogonal subspace is here shown with dashed lines. It is clear that the closest angle after projection, is here with **p1** and not p2.

**Example**: Three Dimensional, so T = 3. Also three vectors, so *p = 3*. Vector w, not matrix w.

X\_1 = (1, 1, 1) / sqrt(3).

X\_2 = (0, 1, 1) / sqrt(2).

X\_3 = (0, 0, 1) / sqrt(1).

w = (0.2, 0.5, 0.8).

Y = X w ~ (0.12, 0.47, 1.27).

Iteration 1: OMP residual: [1.07 1.229 1.269] => x3. OLS MSEs: [0.84 0.58 0.48] => x3.

Iteration 2: OMP residual: [0.337 0.332 0. ]. => **x1**. OLS MSEs: [0.25 0.12 \_\_\_ ] => **x2.**

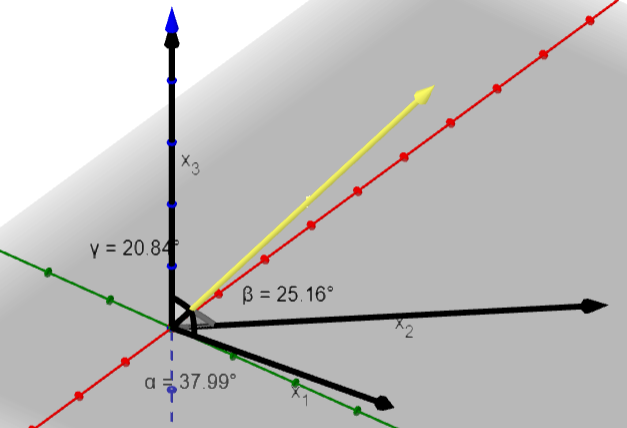
Iteration 3: OMP residual: [0. 0.125 0. ] => **x2**. OLS MSEs: [0. \_\_\_ \_\_\_] => **x1**.

Order is different, as we can see, so both methods are not exactly equal, although they may seem so quite often.

**Geometrix Explanation.**

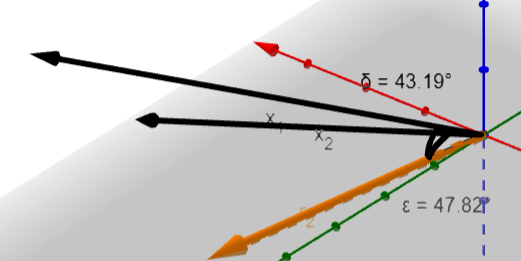
Accompanying GeoGebra: [Link](https://www.geogebra.org/3d/amjnn7pk)

**Step 1:** Pick the atom(vector) that is most correlated with the residual. In other words, pick the atom that has the smallest angle with the current residual.



Remark: We see that the current residual is now orthogonal to the selected atom.

**Step 2**: Pick the atom that is most correlated with the new residual.



**Step 3:** Pick the atom most correlated with the new residual. Only one left.

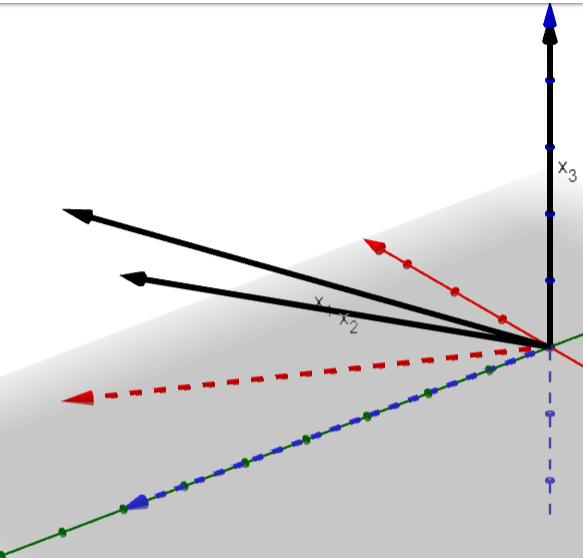
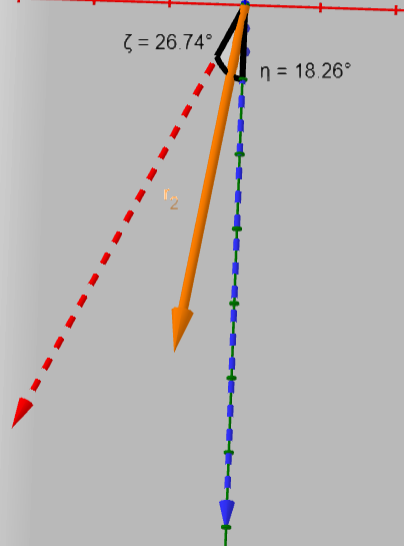
**Orthogonal Least Squares**

Orthogonal Least Squares picks the atom that yields the largest decrease in mean squared error. From a geometric perspective, it picks the atom that has the *smallest angle* with the current residual, *after projecting* the atom on the orthogonal subspace.

**Projecting on Orthogonal Subspace.**

We project the remaining atoms on the orthogonal subspace, and look at their orthogonal projections. We then look at the angles between the residual and the *orthogonal projection* of X.

We have first picked x\_2, the first step of OMP and GLS is always the same. Now, we first project x\_1 and x\_3 onto x\_2.Then, we calculate the angle between the residual and the orthogonal projections of x\_1 and x\_3. Now, we see that x\_1 has the smallest angle, so we pick x\_1.



On the other hand, we picked x\_1, **we now pick x2**.

**Example**: Three Dimensional, so T = 3. Also three vectors, so *p = 3*. Vector w, not matrix w.

X\_1 = (1, 1, 1) / sqrt(3).

X\_2 = (0, 1, 1) / sqrt(2).

X\_3 = (0, 0, 1) / sqrt(1).

w = (0.2, 0.5, 0.25).

Y = (0.2, 0.7, 0.95).

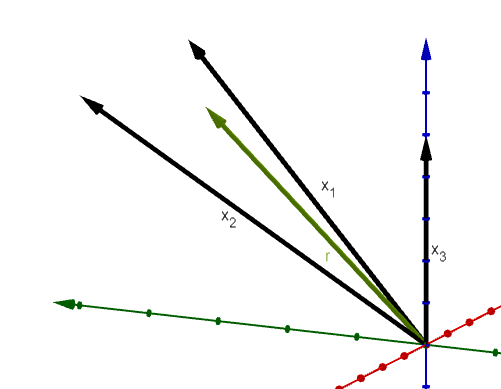
Iteration 1: OMP residual: [1.068 1.167 0.95 ] => x2. OLS MSEs: [0.54 0.27 0.73] => x2.

Iteration 2: OMP residual: [0.115 0. 0.125] => **x3**. OLS MSEs: [0.18 \_\_\_ 0.2 ] => **x1**

Iteration 3: OMP residual: [0.115 0. 0. ] => **x1**. OLS MSEs: [0. \_\_\_ \_\_\_] => **x3**.

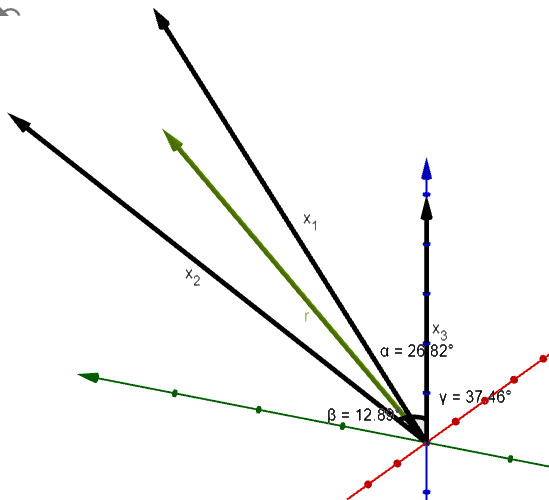
Order is different, as we can see, so both methods are not exactly equal, although they may seem so quite often.

**Geometrix Explanation.**



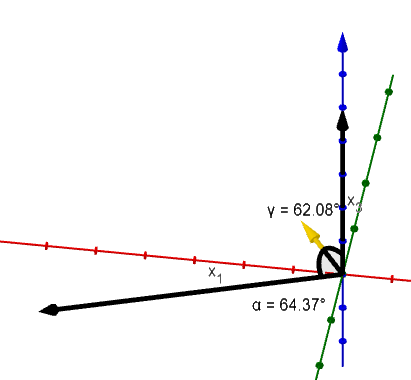
Accompanying GeoGebra: [Link](https://www.geogebra.org/3d/amjnn7pk)

**Step 1:** Pick the atom(vector) that is most correlated with the residual. In other words, pick the atom that has the smallest angle with the current residual.

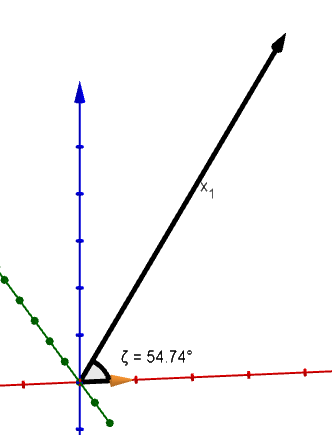


Remark: We see that the current residual is now orthogonal to the selected atom.

**Step 2**: Pick the atom that is most correlated with the new residual.



**Step 3:** Pick the atom most correlated with the new residual. Only one left.



**Orthogonal Least Squares**

Orthogonal Least Squares picks the atom that yields the largest decrease in mean squared error. From a geometric perspective, it picks the atom that has the *smallest angle* with the current residual, *after projecting* the atom on the orthogonal subspace.

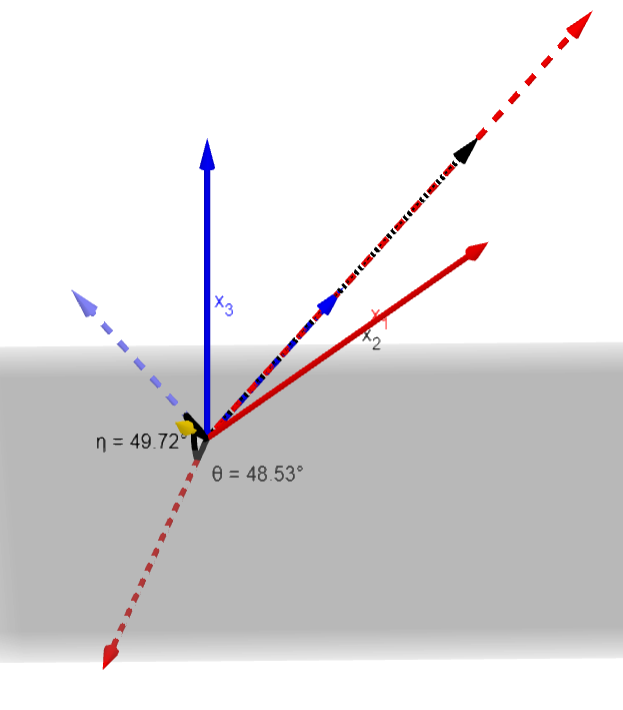
**Projecting on Orthogonal Subspace.**

We project the remaining atoms on the orthogonal subspace, and look at their orthogonal projections. We then look at the angles between the residual and the *orthogonal projection* of X.

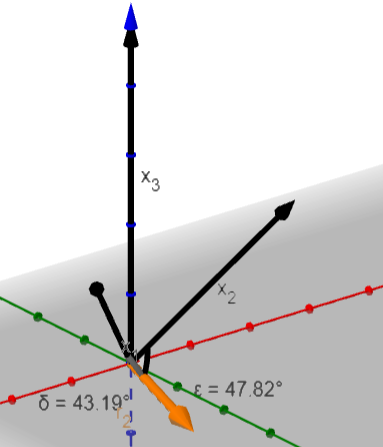
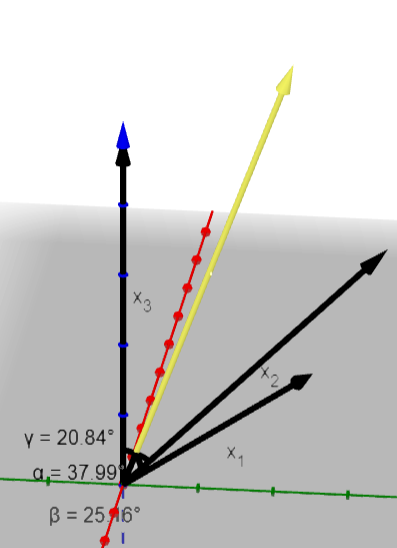
*Crucially, we also need to re-nomalize the edges.*

We have first picked x\_2, the first step of OMP and GLS is always the same. Now, we first project x\_1 and x\_3 onto x\_2.Then, we calculate the angle between the residual and the orthogonal projections of x\_1 and x\_3. Now, we see that x\_1 has the smallest angle, so we pick x\_1.

On the other hand, we picked x\_1.



**Example 2**

****

# Rui Discussion

Would be good to write down a sort of “proposal” for the next 20 weeks. What would we like to achieve in these 20 weeks, in which direction should be go?

More focus on the tuning of regularization parameters of the methodologies we have.

Would like to give some statistical guarantees. For example, correct support recovery of OMP. Another example, correct recovery in the long run when the number of dimensions is large enough.