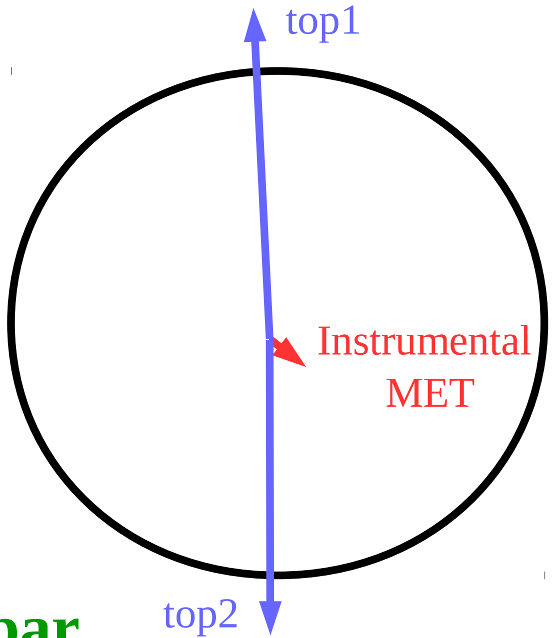
Dark matter + tt in the dilepton channel: status

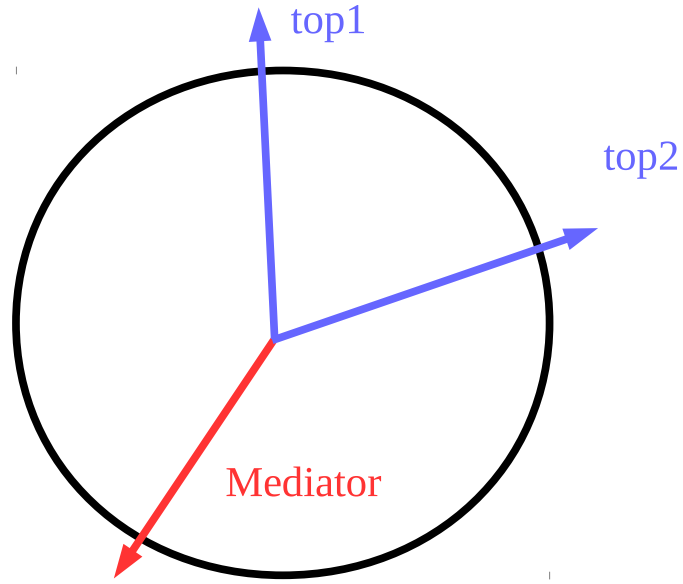
P. Martinez

Top reconstruction

- The top reconstruction is an attempt to estimate the pt of the tops in a ttbar-like event
- Why is this interesting? → Consider the quantity -(**pt**(top_1) + **pt**(top_2))
  - In a pure ttbar event the tops should be balanced and this vector should be very small
  - In a ttbar+DM event however this corresponds to the **pt** of the mediator



ttbar



ttbar+DM

How can we reconstruct the tops?

Forgetting the mediator for now...

- One could think it's not possible to do such a thing because we don't know the neutrinos
  - The 3 components of the momenta of the neutrinos are unknown → **6 unknowns**
- However in a ttbar event there are things we know:
  - The momentum of the 2 b-jets and the mometum of the 2 leptons → 12 variables
  - We know that the lepton and neutrino come from a W
    - M(nu_1 + l_1) = M_W and M(nu_2 + l_2) = M_W → **2 equations**
  - We also know that the W and the b-jet come from a top
    - M(nu_1 + l_1 + b_1) = M_top and M(nu_2 + l_2 + b_2) = M_top → **2 equations**
  - Finally we know that the nuetrinos add up to give the MET
    - nu_1_x + nu_2_x = MET_x and nu_1_y + nu_2_y = MET_y → **2 equations**
- The system gives a 4th order polynomial, p(nu) on the x component of one of the neutrinos

But life's always tougher than expected

- Indeed even when we say we know it's not so clear that we really know
  - We don't know how to assign lepton_1 to b_jet_1, and lepton_2 to b_jet_2
  - Moreover → we could have more than 2 jets in the event, which is the right combination?
  - And what about the masses of the top and the W? They are not really fixed (it's a BW!)
  - MET is a variable which is not measured with great precision. What if it's wrong?
  - Even if we found everything right above p(nu) can give up to 4 physical solutions
- This has a few implications that we should think about:
  - In one particular ttbar event we might find no solutions at all because:
    - We don't have the right combination of jets/leptons
    - We have a weird value of m(W) or m(top), or the MET it's not right
  - But we can also find many solutions corresponding to different combinations

Solutions are coming from the top group

- First of all: we need to think about the top reconstruction as an statistical problem
  - It's not giving the momentum of the top, it's giving a “likely” momentum of the top
- What we do is to try to reconstruct the top with different combinations of leptons and jets
  - If there are 2 b-jets in the event we try with those
  - If not, or they didn't give any solution, we try with the 1 b-jet and all the others
- Then for every combination of lepton-jets we randomly sample on 2 things:
  - The jet energy scale of the jets considered → gaussian with jet resolution
  - The masses of the W and the top → Breit-Wigner with the 2 corresponding widths
- For each we solve the 4th order polynomial → and then what
  - We take the solution that makes the mass of the top-antitop system smaller
  - We define a weight based on the true m_lb shape of this combination

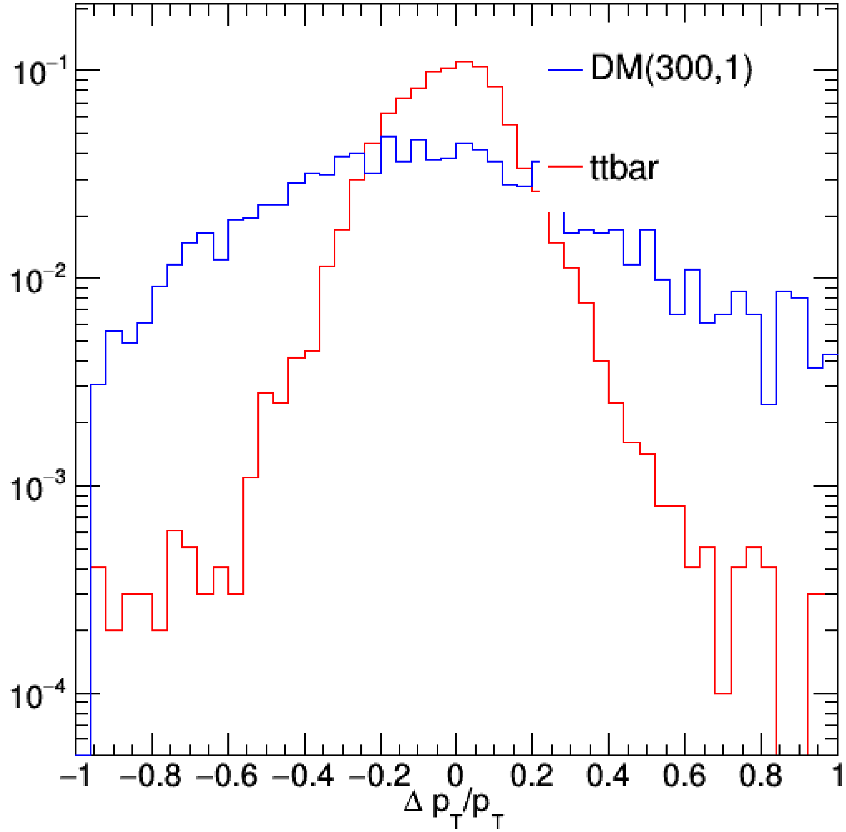
How does it work?: Some pseudocode

```
[weight, top1, top2] = solve(l1, l2, jets, MET)
lweight_max = -1; top1 = 0; top2 = 0;
For j in all combinations of leptons and jets:
    [lweight, ltop1, ltop2] = solveIt(l1, l2, jet1, jet2, MET)
    if(lweight > lweight_max)
        lweight_max = lweight
        top1 = ltop1; top2 = ltop2
Return lweight_max, top1, top2
```

```
[weight, top1, top2] = solveIt(l1, l2, jet1, jet2, MET)
weight = 0; top1 = 0; top2 = 0;
For j in Number:
    [j1, j2, massesW, massesT] = smearTheJetsAndMasses()
    [tops1, tops2] = getSolutions(j1, j2, massesW, massesT)
    [ltop1, ltop2] = selectSmallerTopSystemMass(tops1, tops2)
    [lweight] = probabilityOf(mlb1, mlb2)
    Weight = weight + lweight
    Top1 = top1 + lweight*ltop1; top2 = top2 + lweight*ltop2
top1 = top1/weight; top2 = top2/weight
Return weight, top1, top2
```

What do we get?

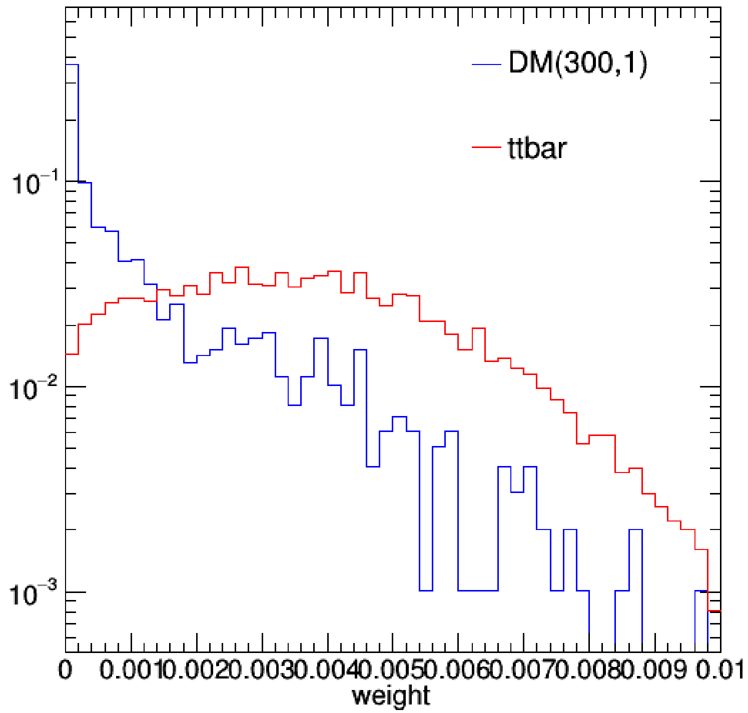
- When appling this method to ttbar events we reconstruct 99% of events with 20% resolution
- But when applying to DM we only reconstruct 20% of the events with very bad resolution

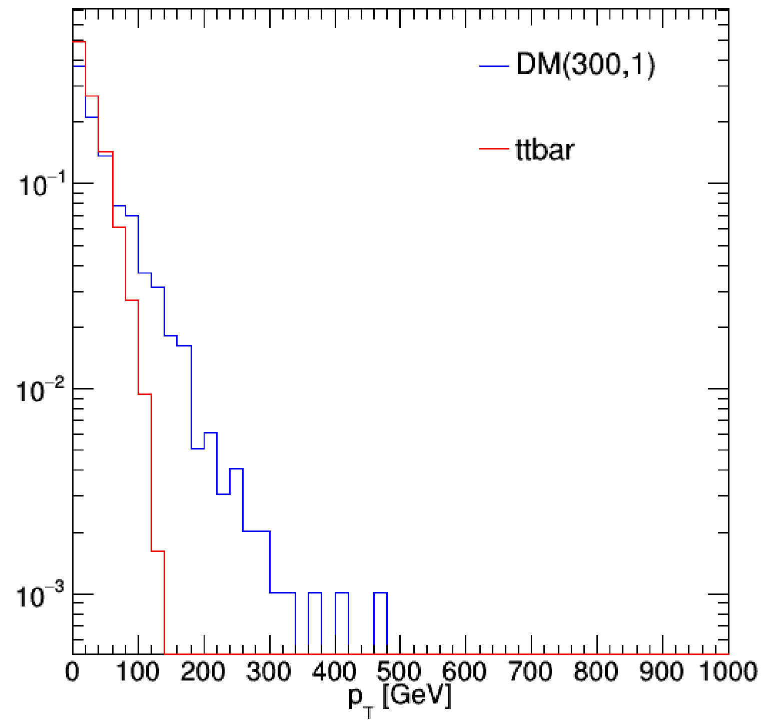


Why this behaviour?

- It's clear: in the DM case MET is not anymore originating from neutrinos but also the DM
- The constraints we are imposing to our ttbar topology are not correct anymore
- But how it's possible that we still see solutions?
  - Simple. We are smearing the jets and trying different combinations
  - Sometimes the MET matches a combination of jets + smearing + W, top masses
- When this is the case the reconstructed top and the weights are not so good as for ttbar
- In particular the top_pts will tend to be higher since they will be accomodating extra MET
- But this is very good news because we can use them as a discriminator!
- However we will need to deal with the low efficiency of reconstructed events (for DM)

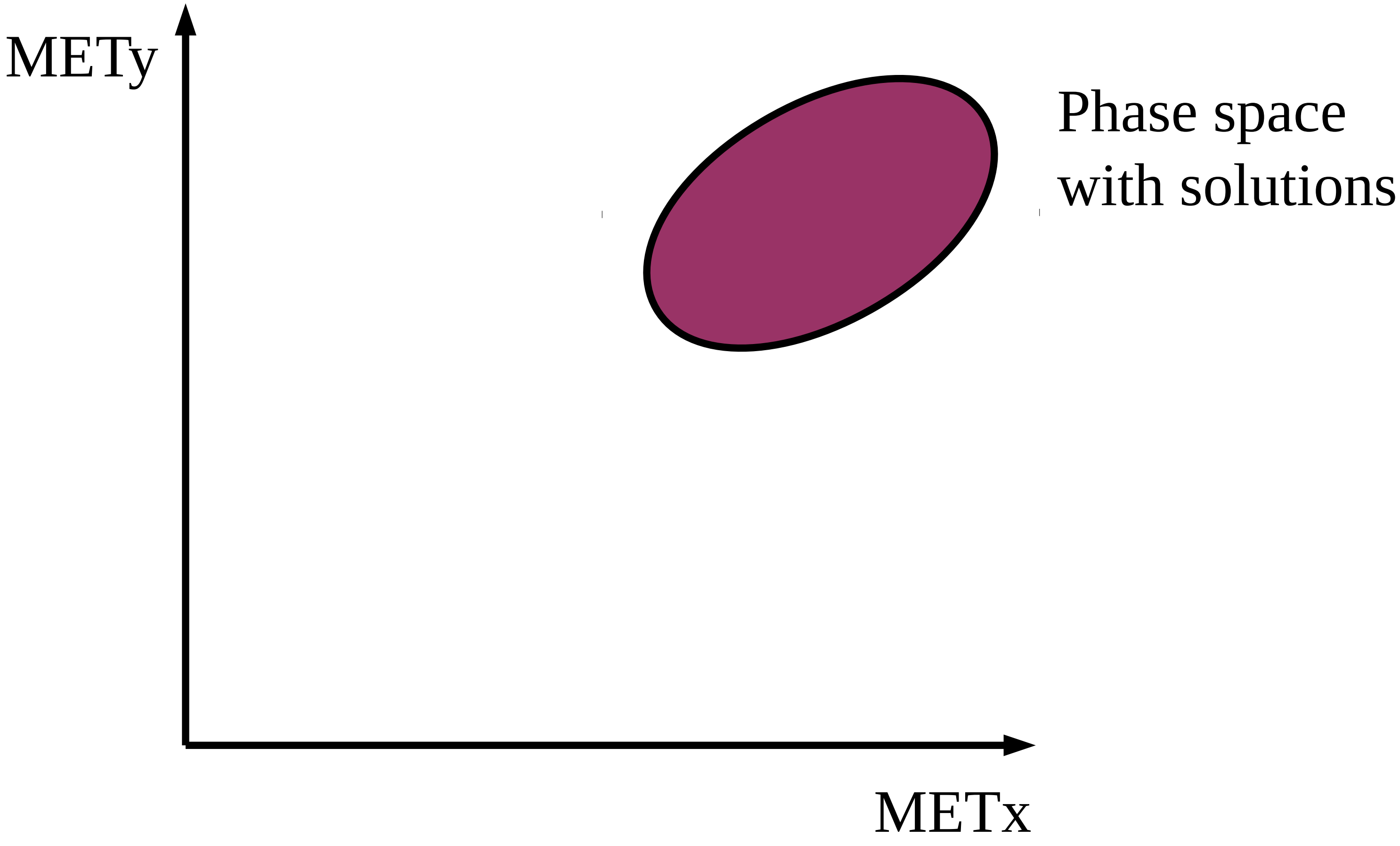
Discrimination





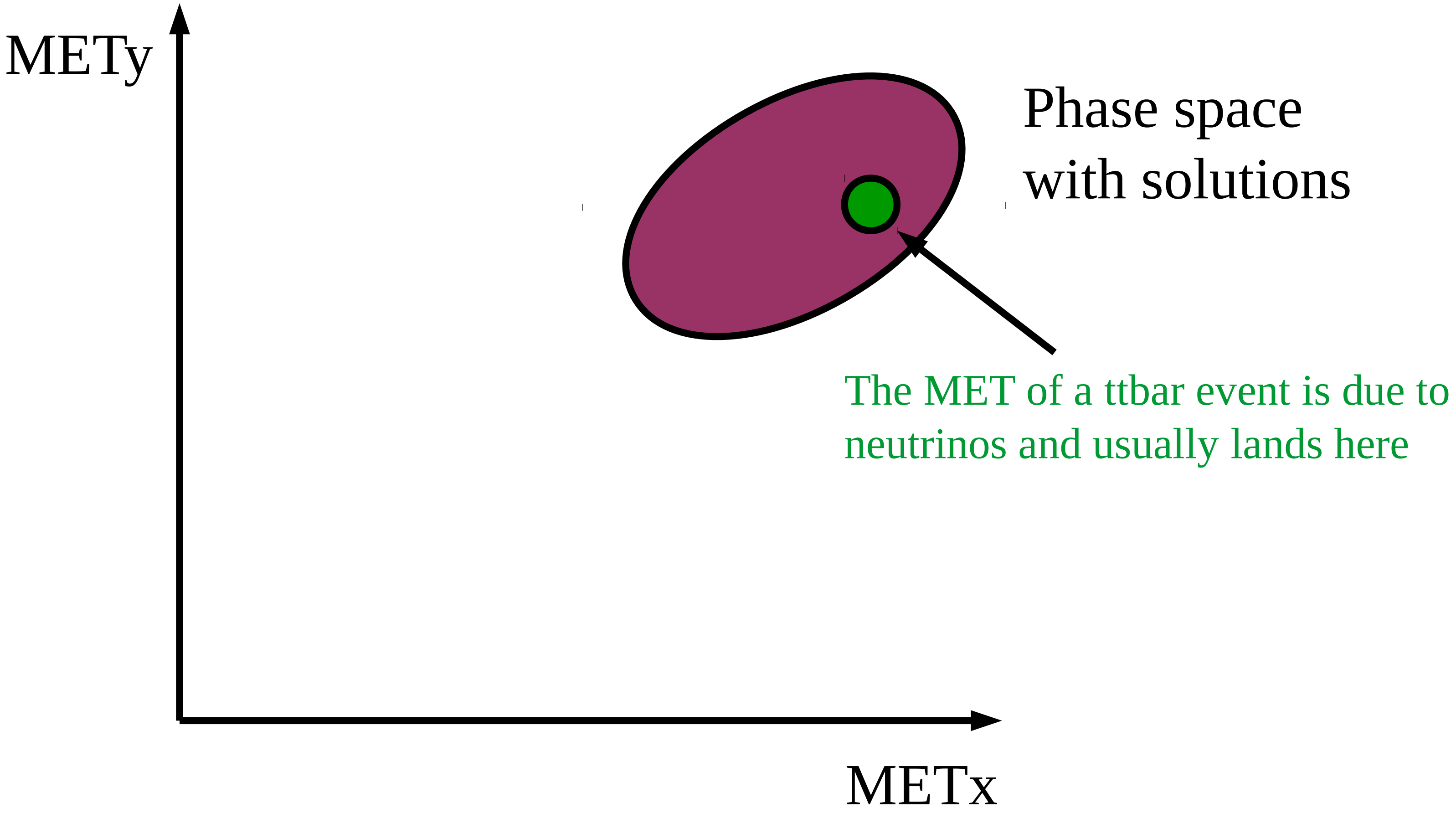
Can we correct the efficiency problem?

- As we have seen by applying this method we would be throwing away 75% of DM events
- The value of MET is very far away for a place giving real solutions (smearing not enough)



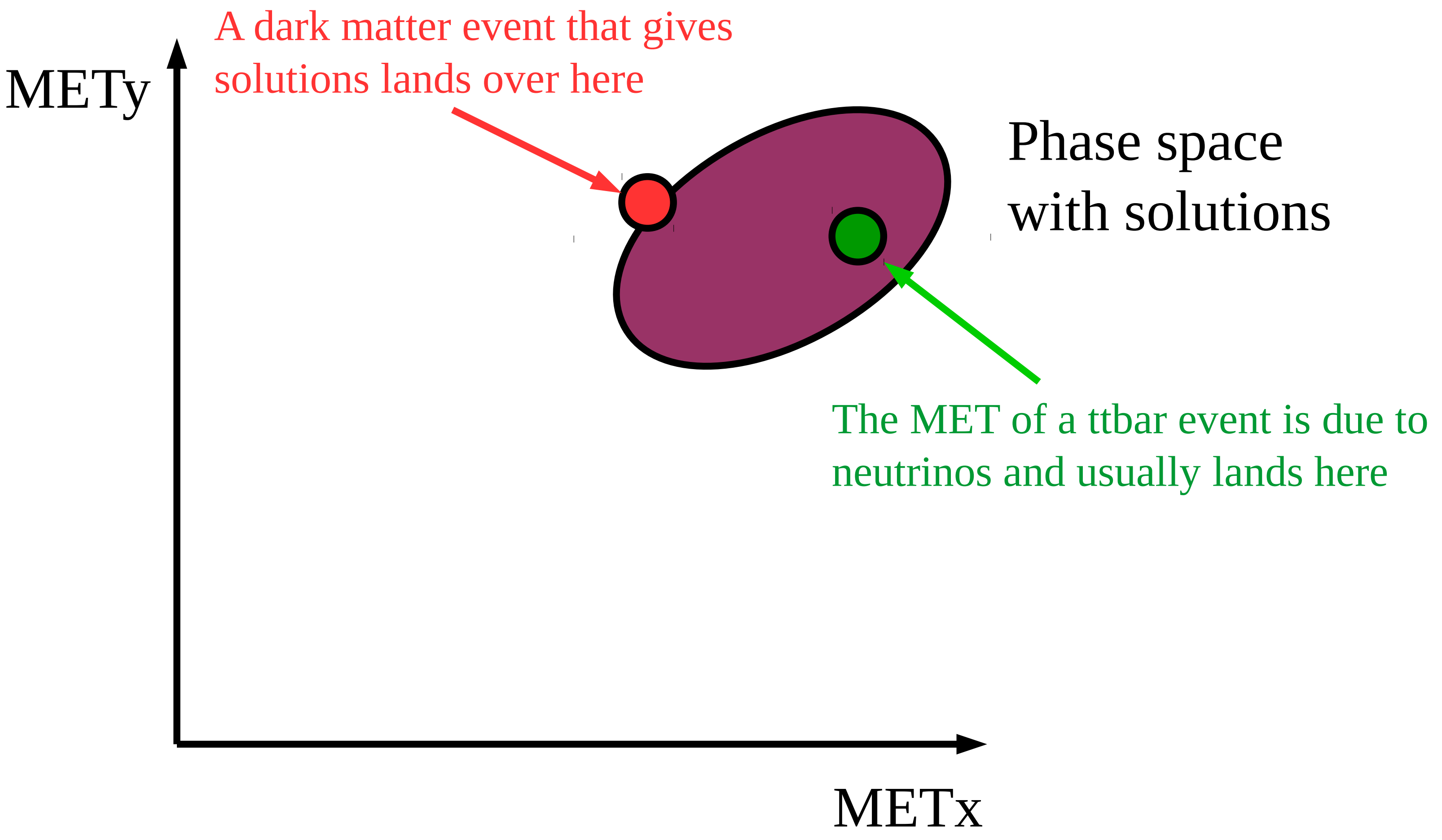
Can we correct the efficiency problem?

- As we have seen by applying this method we would be throwing away 75% of DM events
- The value of MET is very far away for a place giving real solutions (smearing not enough)



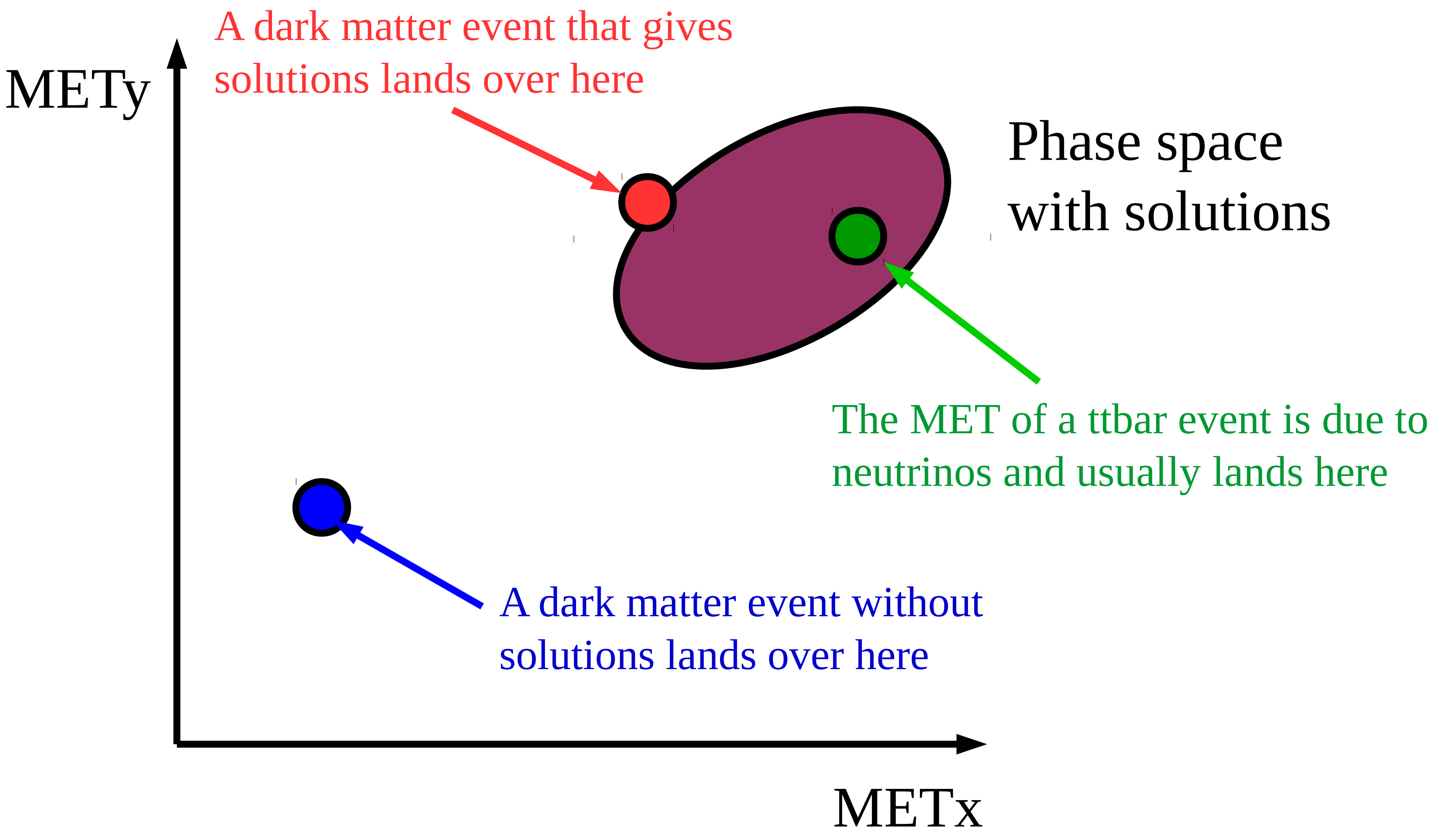
Can we correct the efficiency problem?

- As we have seen by applying this method we would be throwing away 75% of DM events
- The value of MET is very far away for a place giving real solutions (smearing not enough)



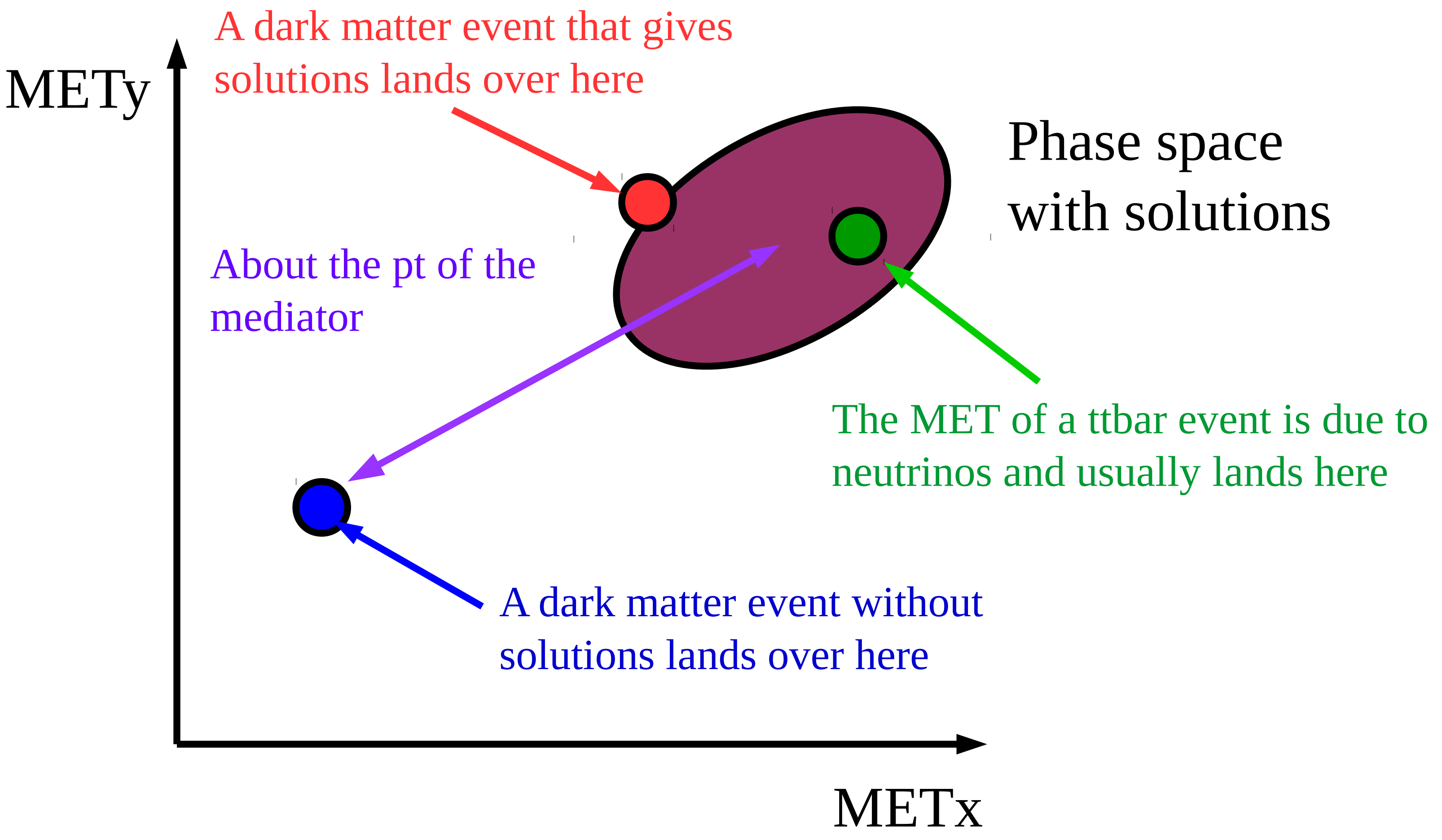
Can we correct the efficiency problem?

- As we have seen by applying this method we would be throwing away 75% of DM events
- The value of MET is very far away for a place giving real solutions (smearing not enough)



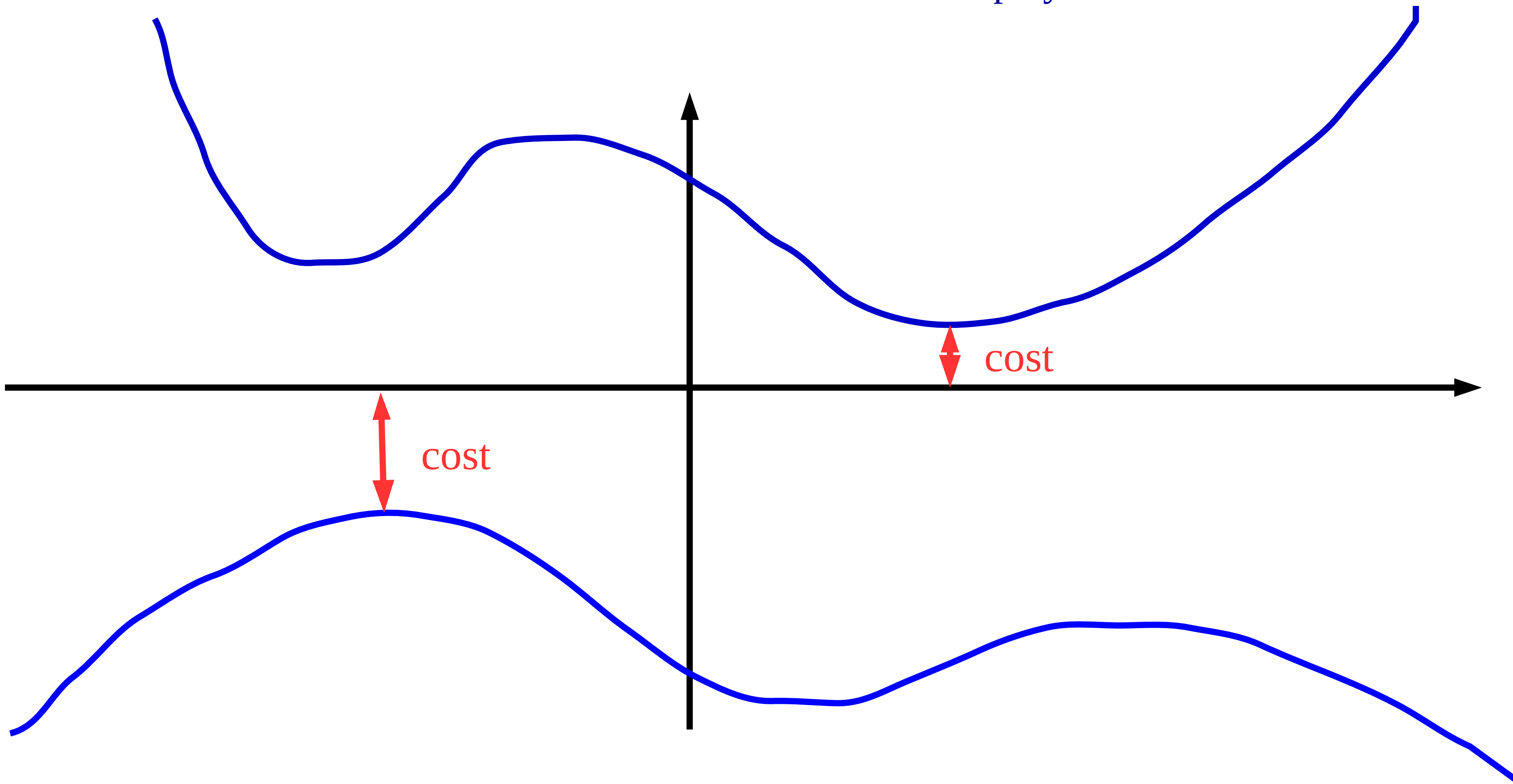
Can we correct the efficiency problem?

- As we have seen by applying this method we would be throwing away 75% of DM events
- The value of MET is very far away for a place giving real solutions (smearing not enough)



To go there we can define a metric

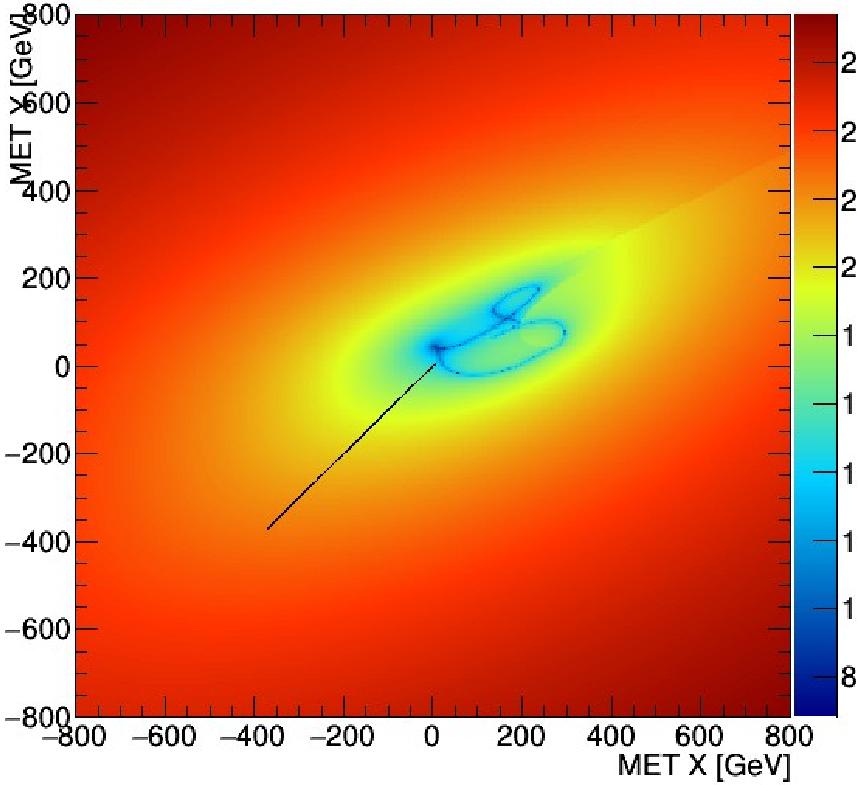
- When there are no solutions it means the polynomial is always above or below the x axis
- We can define the “cost” as the minimal distance between the polynomial and the axis

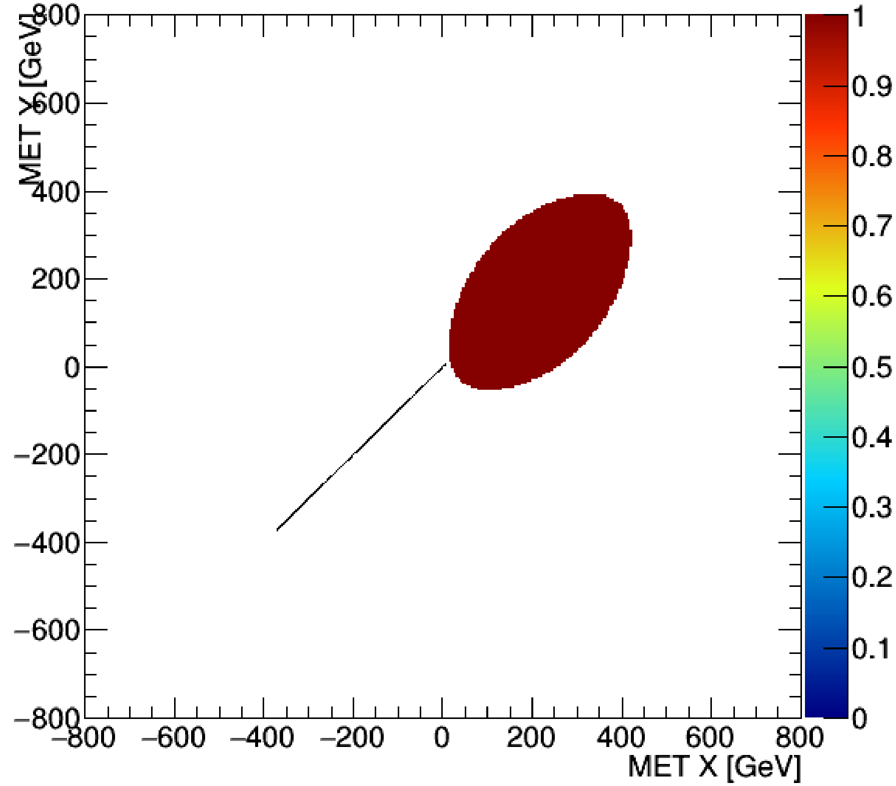


Gradient descent

- Once we have this metric can do an iterative gradient descent to get into the solution space
- How? We estimate using the definition of derivative:
- We make most of DM events converging points

$$-\nabla \vec{cost} = \begin{pmatrix} \frac{\partial\, cost}{\partial\, MET_x} \\ \frac{\partial\, cost}{\partial\, MET_y} \end{pmatrix}$$





Where are we?

- If we assume that we have to give a status report on the 7th of December we have 6 weeks

Week 1 (the technical week)

- Synchronize framework between Cedric and Juan
- Ensure we can produce plots in “minutes” (minitrees)
- Synchronize baseline selection
- Summary of new variables inspected by Cedrid
- Neural network and top reconstruction in place
- Meeting on Friday to see the checklist

DONE

Partially DONE

Need to check S/sqrt(B) for top reconstruction

Code and ideas is 95% in place

Week 2 (reference + top reconstruction)

- Check the S/sqrt(B) with cuts in DeltaPhi and MET
- Check the S/sqrt(B) adding also Cedric's cuts
- Work in the top reconstruction algorithm
- Check the S/sqrt(B) adding top-based variables
- Meeting on Friday to see the checklist
- Decide which variables are the most useful

Timeline II

- If we assume that we have to give a status report on the 7th of December we have 6 weeks

Week 3 (data + the neural network)

- Start doing detailed comparisons data/MC
- Take the best ranked variables and apply the ANN
- Evaluate S/sqrt(B) of the final discriminator
- Meeting on Friday to see the checklist
- Decision: Several regions for several points?

MISSING

MISSING

Week 4 (background prediction)

- Revive the background prediction methods
- Perform CLOSURE TEST in MC
- Check systematics in detail
- Meeting on Friday to see the checklist
- Decide which variables are the most useful

This week

Timeline II

- If we assume that we have to give a status report on the 7th of December we have 6 weeks

Week 5 (Interpretation + AN)

- Start writing an Analysis Note with partial results
- Setup all the code for setting upper limits
- Estimation of systematic uncertainties in signal
- Ask for status report to the EXO conveners

Week 6 (contingency)