SPONSORED BY THE

Federal Ministry of Education and Research

Deutsche Forschungsgemeinschaft

DFG

SFB 676 – Projekt B2

# Extended Fits with Many Parameters
such as
# Constrained Kinematic Fits

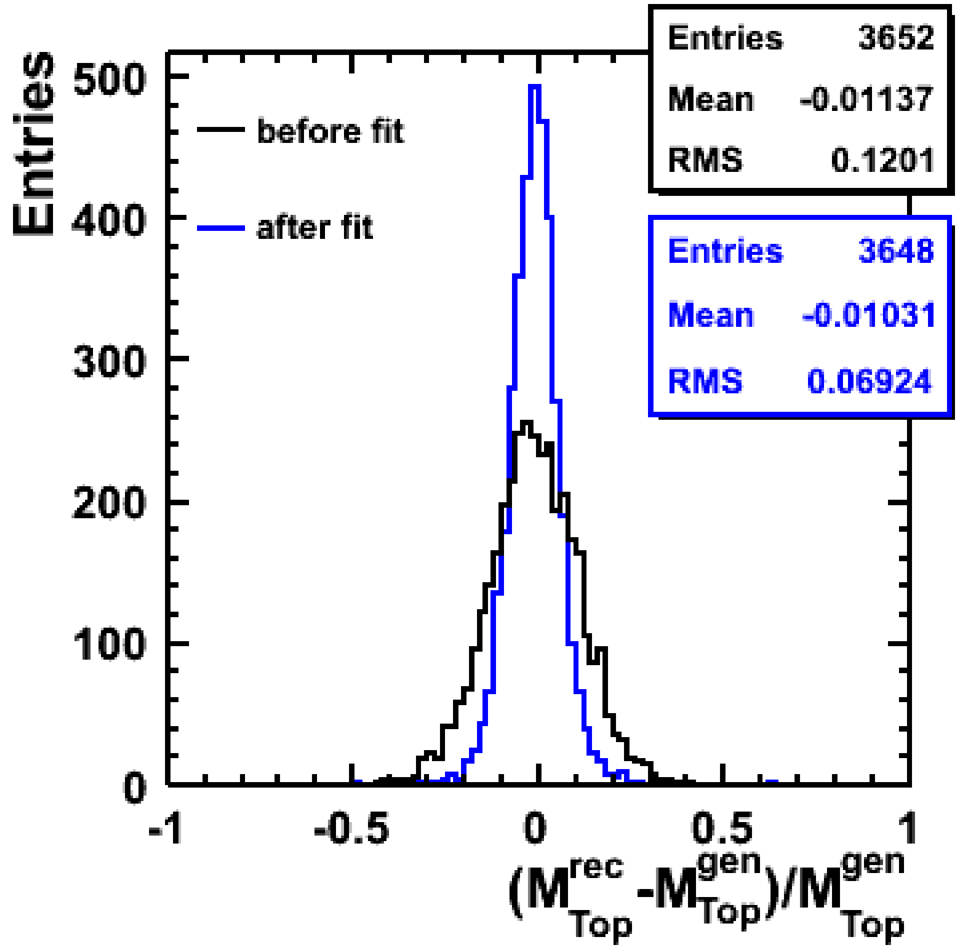
B. List, B. Mura, C. Sander (*Universität Hamburg*)

HGF Analysis Centre Statistics School / Mainz / 4th-7th April 2011

PHYSICS AT THE TERA SCALE
Helmholtz Alliance

# Outline

- *Reminder:* least square fits
- Constrained fits with technique of Lagrangian Multipliers
- Linearization of minimization problems
- Kinematic fits:
    - mass constraints
    - momentum balance
    - unmeasured parameters
- Iterative algorithm and optimization
- Alternative method: minimization of cost function




**Example:** Improvement of top mass resolution using a kinematic fit

# Reminder: Method of Least Squares

$N$ **measurements**: $y_i$ with **variances** $\sigma_i^2$

In case of statistically independent variables the **covariance matrix** is diagonal

$$V[y] = \begin{pmatrix} \sigma_1^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & 0 & \cdots & 0 \\ & \cdots & & & \\ 0 & 0 & 0 & \cdots & \sigma_N^2 \end{pmatrix} \quad V^{-1}[y] = \begin{pmatrix} 1/\sigma_1^2 & 0 & 0 & \cdots & 0 \\ 0 & 1/\sigma_2^2 & 0 & \cdots & 0 \\ & \cdots & & & \\ 0 & 0 & 0 & \cdots & 1/\sigma_N^2 \end{pmatrix}$$

**Model** $f_i(a_1 \ldots a_M)$ with $M$ **free parameters** $a_j$

Normalized **residuals**

$$\chi^2 = \sum_{i=1}^{N} \frac{r_i^2}{\sigma_i^2} = \sum_{i=1}^{N} \frac{(y_i - f_i(\vec{a}))^2}{\sigma_i^2} = \sum_{i,j=1}^{N} (y_i - f_i(\vec{a}))^T V_{ij}^{-1}[y](y_j - f_j(\vec{a}))$$

The best fitting model parameters should minimize this expression

# Reminder: Method of Least Squares

Minimization of $\chi^2$: derivatives with respect to all model parameters $a_i$ should vanish

$$\frac{d\chi^2}{da_i} \stackrel{!}{=} 0$$

**Matrix notation** for **linear models**

$$\vec{f} = A \cdot \vec{a}$$

$\rightarrow$ Sum of residuals

$$\chi^2 = \vec{r}^{\,T} V^{-1}[y] \vec{r} = (\vec{y} - A \cdot \vec{a})^T V^{-1}[y] (\vec{y} - A \cdot \vec{a})$$

...

**Normal equation:** $(A^T V^{-1}[y] A) \cdot \vec{a} = A^T V^{-1}[y] \vec{y}$

**Solution:**

$$\begin{aligned} \vec{a}_{\text{estimate}} &= (A^T V^{-1}[y] A)^{-1} A^T V^{-1}[y] \vec{y} \\ V[\vec{a}_{\text{estimate}}] &= (A^T V^{-1}[y] A)^{-1} \end{aligned}$$

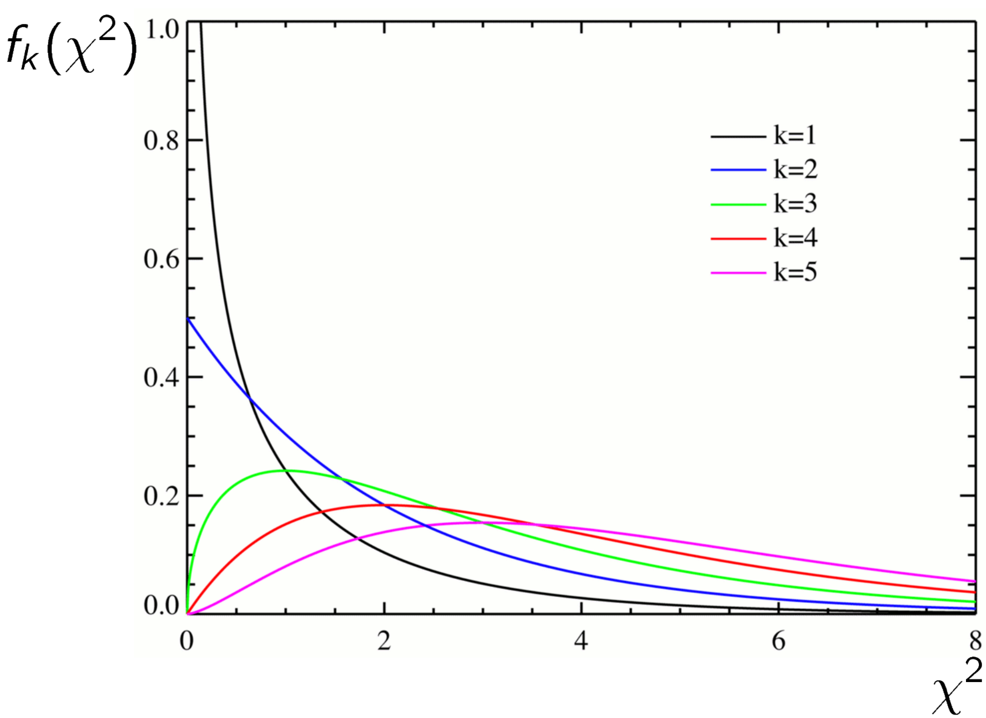
# $\chi^2$ Distribution & Degrees of Freedom

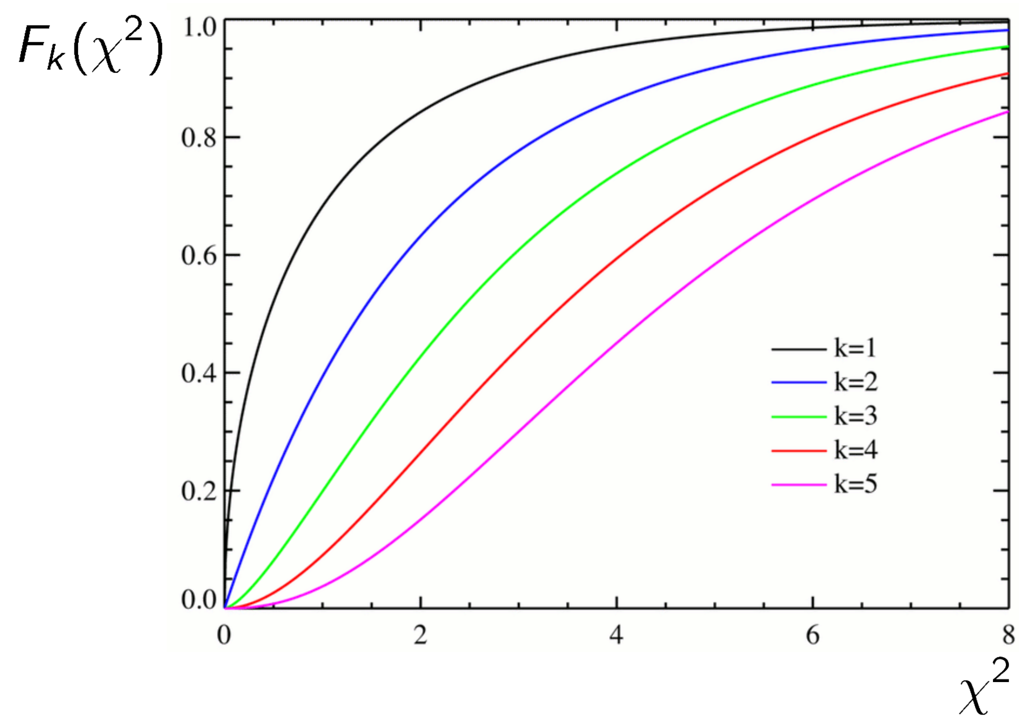
For $N$ measurements with $M$ fitted parameters,

$$\chi^2 = \sum_{i=1}^{N} \frac{r_i^2}{\sigma_i^2} = \sum_{i=1}^{N} \frac{(y_i - f_i(\vec{a}))^2}{\sigma_i^2}$$

is distributed according to $\chi^2$ distribution with $k = N - M$ degrees of freedom.

If we add $P$ **constraints** between fitted parameters (see later) → number of free parameters is reduced by $P$ and the d.o.f. are increased: $\boldsymbol{k = N - M + P}$



**Pulls:** normalized residuals with respect to true values

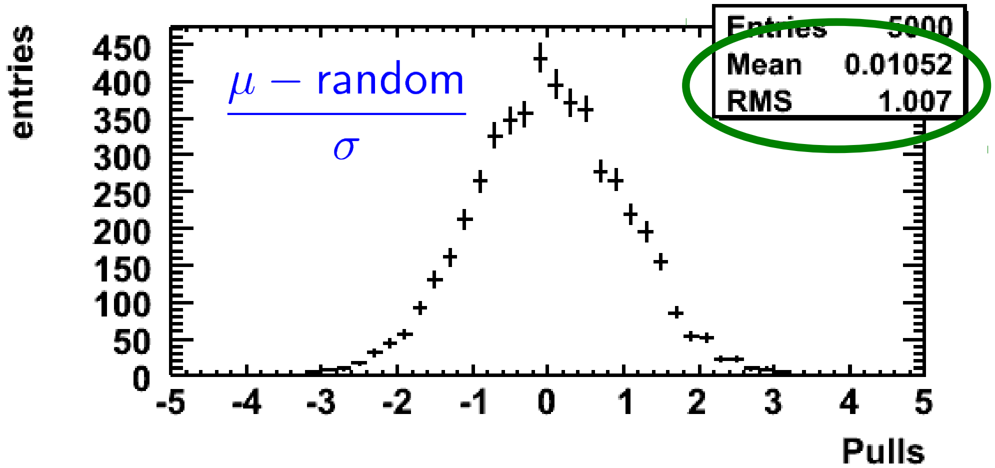
$$\text{pulls} = r_i = \frac{y_i^{\text{fit}} - y_i^{\text{meas}}}{\sigma_i^{\text{fit}}}$$

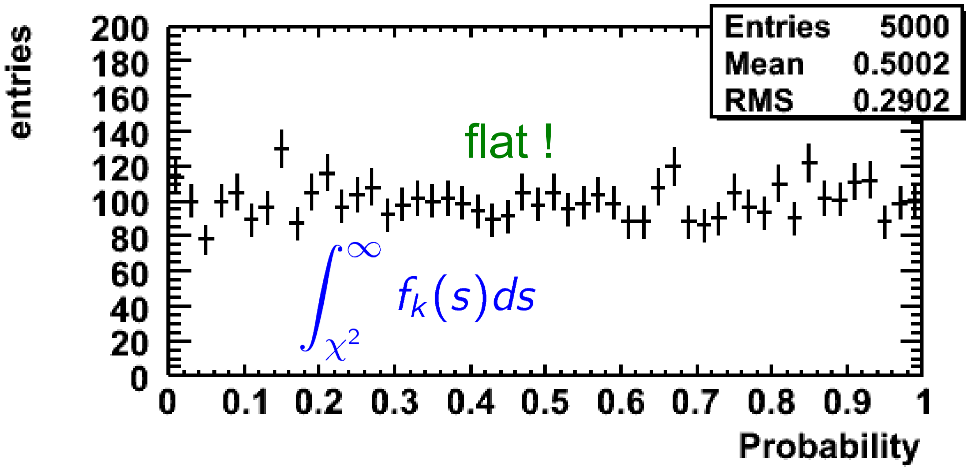
Properties of pulls for correct model:

- normal distributed with mean $\mu_r = 0$. If not → systematic errors of measurements.
- variance $\sigma_r = 1$. If large than 1 → errors are underestimated else overestimated.

If $\chi^2$ follows a $\chi^2$-distribution with $k$ degrees of freedom → corresponding probability distribution is flat!

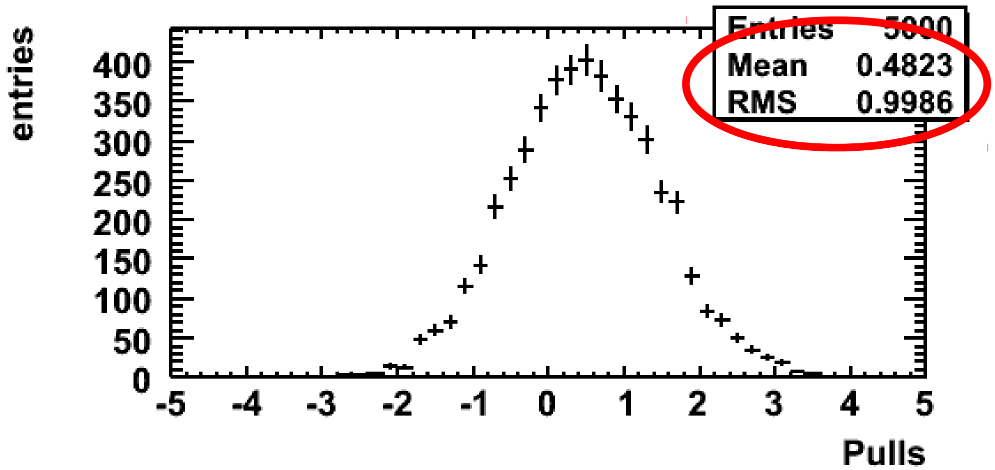
**Example:** normal distribution ( $\mu$ = 1 and $\sigma$ = 1) of random variable (one measurement $N$ = 1, no fit parameter $M$ = 0 and no constraint $P$ = 0) → $k$ = 1

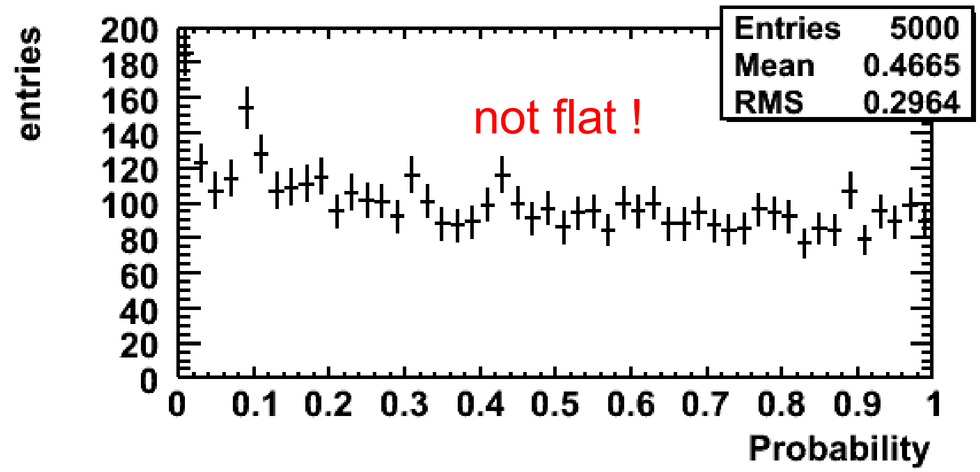


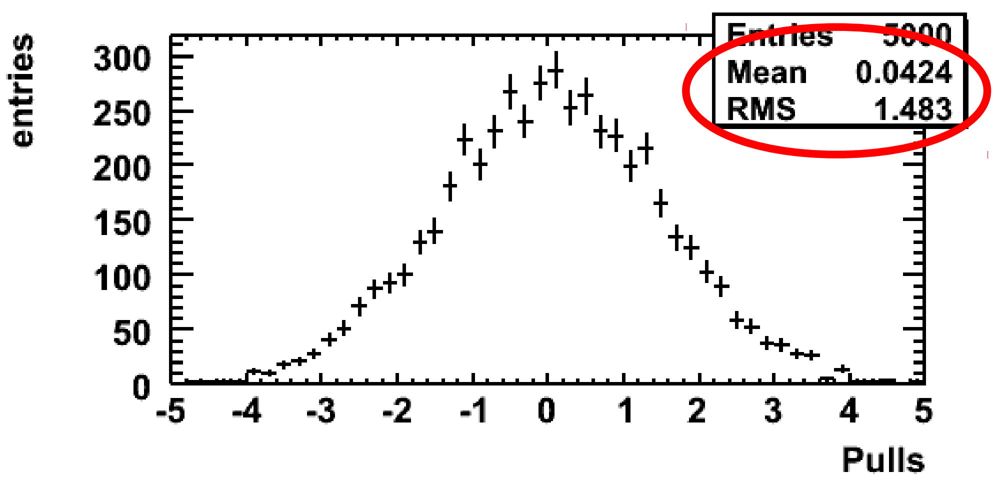
# Example: Interpretation of Results

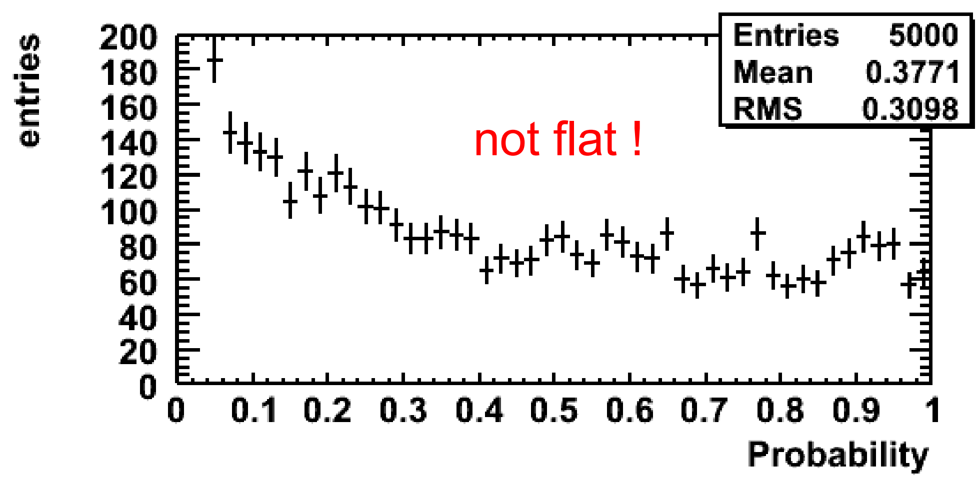
**Incorrect model I:** true value $\mu$ offset by +0.5 $\sigma$ w.r.t. model
→ pulls shifted to the right, variance OK, probability rises at low values




**Incorrect model II:** true error $\sigma$ larger by factor 1.5 w.r.t. model
→ mean of pulls = 0, variance larger than 1, probability rises at low values



# One Parameter for Each Measurement

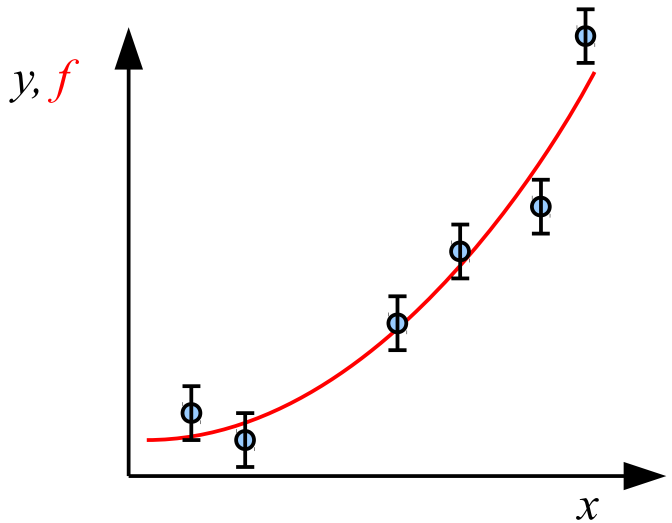
Linear least squares: fit a function with few parameters to many measurements:

**Example:** Quadratic fit

$$f_i(\vec{a}) = a_0 + a_1 \cdot x_i + a_2 \cdot x_i^2 \rightarrow \chi^2 = \sum_{i=1}^{N} \frac{(y_i - f_i(\vec{a}))^2}{\sigma_i^2}$$

Matrix form: $\vec{f} = A \cdot \vec{a}$

$$\begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{pmatrix} = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ & \cdots & \\ 1 & x_N & x_N^2 \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix}$$

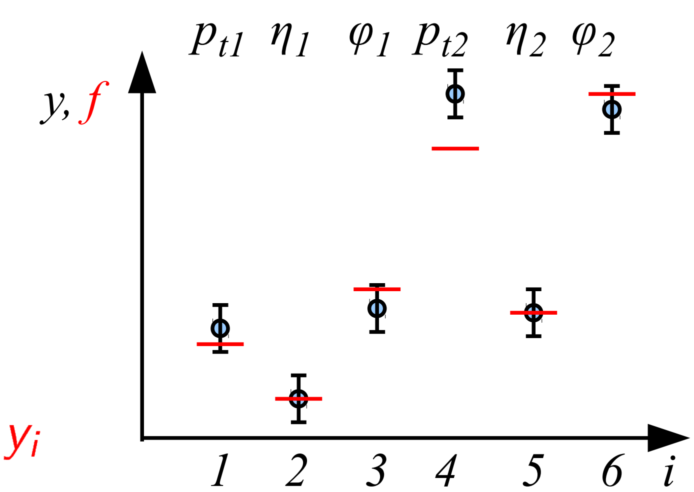



**Now:** One fit parameter for each measurement

$$f_i(\vec{a}) = a_i \rightarrow \chi^2 = \sum_{i=1}^{N} \frac{(y_i - a_i)^2}{\sigma_i^2}$$

Matrix form: $A = I$

$$\begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & \\ 0 & 1 & 0 & \vdots \\ \vdots & & \ddots & \\ & \cdots & & 1 \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_N \end{pmatrix}$$




The solution of the unconstrained problem is trivial: $a_i = y_i$

# Constraints

**Problem:** the model parameters are not entirely free but the model has to fulfill some special condition (constraint)

**Example:**

Two measurements $y_1 = 1$ and $y_2 = 2$ with error $\sigma = 1$

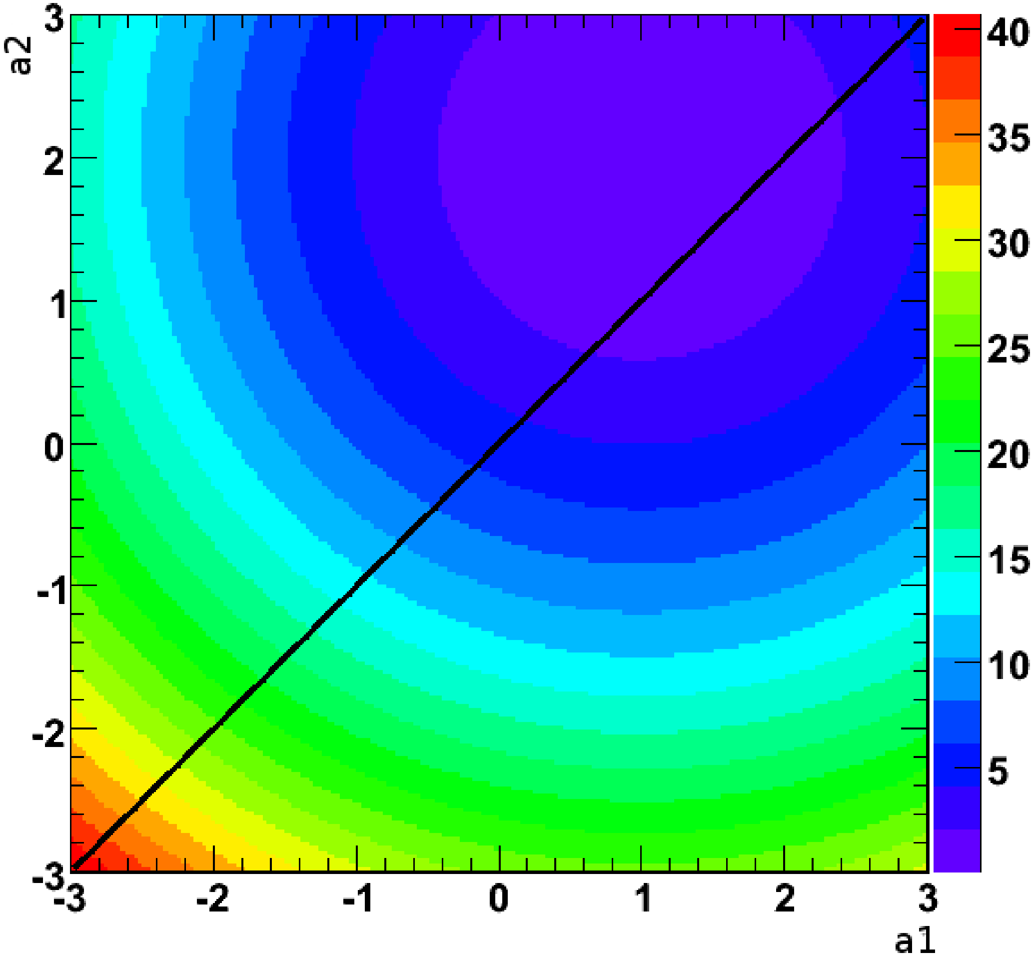
$$\chi^2 = (1 - a_1)^2 + (2 - a_2)^2$$

$\rightarrow$ Minimum at $a_1 = 1$ and $a_2 = 2$

Wanted is the minimum for $a_1 = a_2$

$$a_1 - a_2 = 0$$

A widely used method to solve such a minimization problem with additional constraints is the **Method of Lagrangian Multipliers**



# Method of Lagrangian Multipliers

The first step is to formulate each constraint as equation that equals zero

$$c(\vec{a}) \stackrel{!}{=} 0$$

Definition of a new function $L$ for $P$ constraints

$$L(\vec{y}, \vec{a}) = \chi^2 + 2 \cdot \sum_{i=1}^{P} \lambda_i c_i(\vec{a})$$

with **Lagrangian Multipliers** $\lambda_i$

**Partial derivative w.r.t. $\lambda_i$:**

$$\frac{\partial L}{\partial \lambda_i} = 2 \cdot c_i(\vec{a}) \stackrel{!}{=} 0$$

→ solution must fulfill the constraints!

**Partial derivatives w.r.t. $a_i$:**

$$\frac{\partial L}{\partial a_i} = \frac{\partial \chi^2}{\partial a_i} + 2 \cdot \sum_{j=1}^{P} \lambda_i \cdot \frac{\partial c_j}{\partial a_i} \stackrel{!}{=} 0 \rightarrow \nabla \chi^2 = -2 \cdot \sum_{j=1}^{P} \lambda_j \nabla c_j$$

→ for one constraint ($P$ = 1): gradients of $\chi^2$ function $\nabla\chi^2$ and of constraint function $\nabla c$ must be parallel !

→ $\chi^2$ function has a local minimum on the constraint contour !

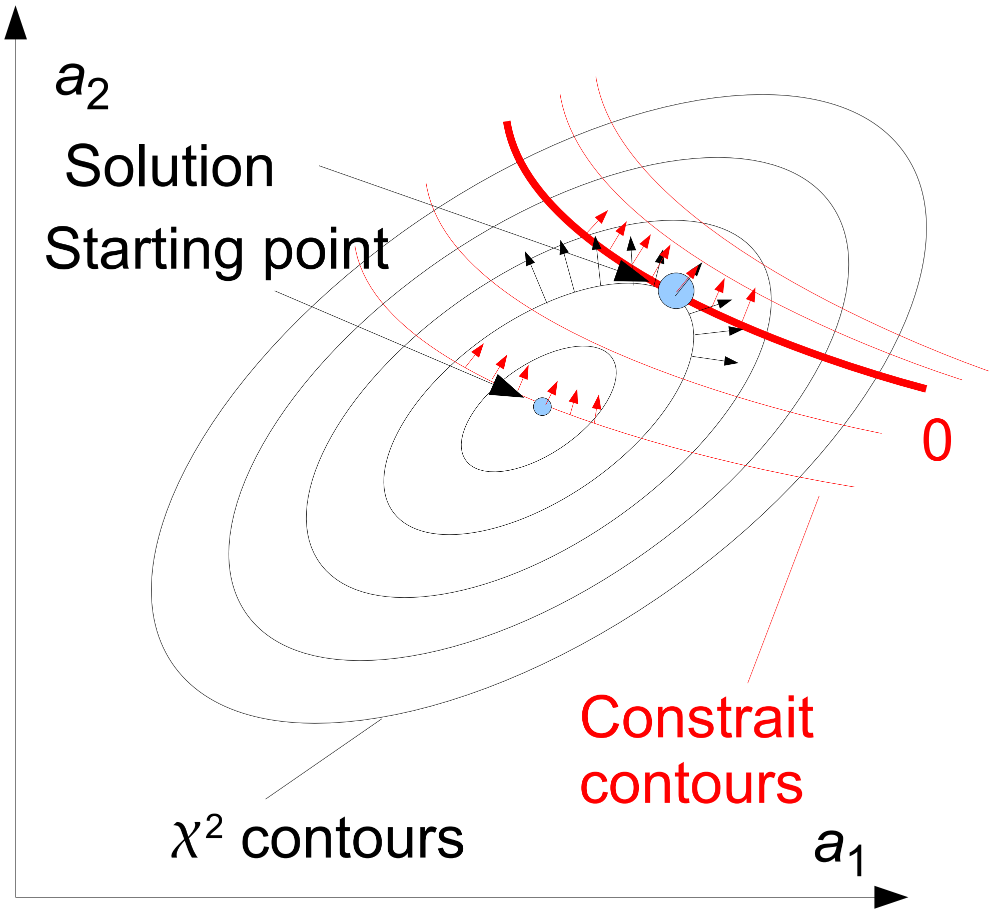
**Visualization in 2D:**

- Solution must lie on **zero contour** of the constraint

$$\frac{\partial L}{\partial \lambda} = 2 \cdot c(\vec{a}) \stackrel{!}{=} 0$$

- Constraint line must be parallel to $\chi^2$ contour at solution, i.e. **gradients** of $\chi^2$ and constraint must be **parallel**

$$\nabla \chi^2 = -2\lambda \nabla c$$




Example for 2 parameters $a_1$, $a_2$ and one constraint

# Explicit Solution of the Example

Back to our problem: minimization of $\chi^2 = (1 - a_1)^2 + (2 - a_2)^2$

subject to the constraint $c(a_1, a_2) = a_1 - a_2 \stackrel{!}{=} 0$

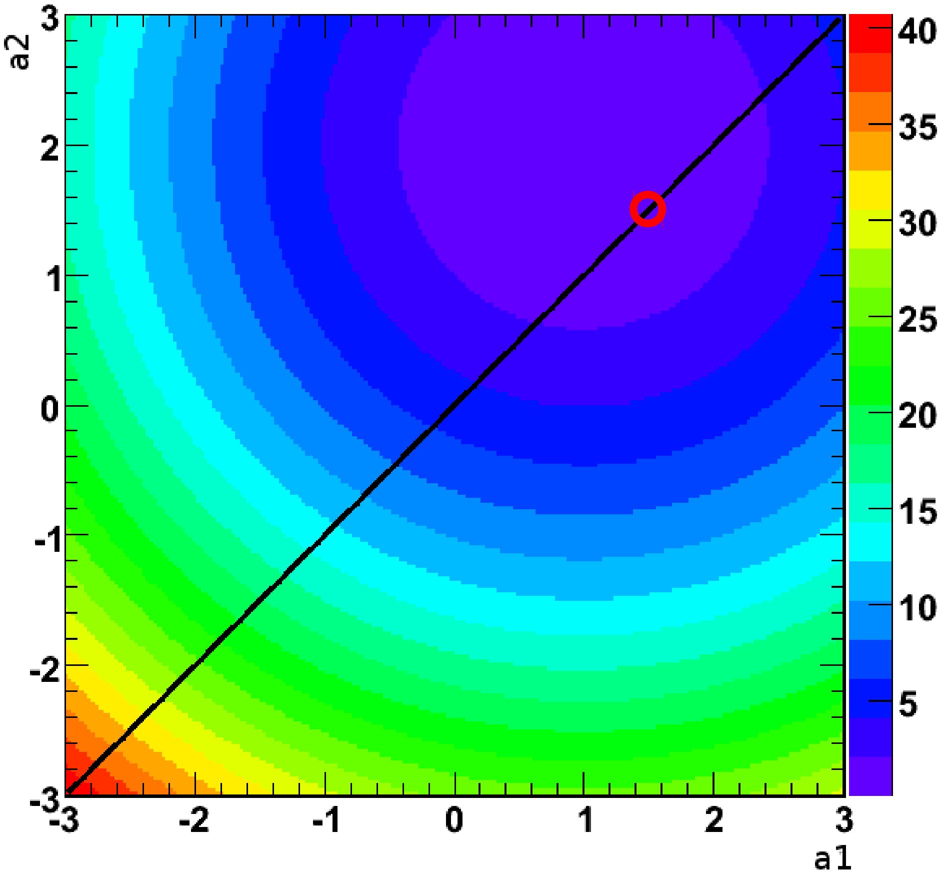
Definition of a Lagrange function $L(a_1, a_2) = (1 - a_1)^2 + (2 - a_2)^2 + 2 \cdot \lambda \cdot (a_1 - a_2)$

Partial derivatives have to vanish:

$$\frac{\partial L}{\partial a_1} = 2 \cdot (a_1 - 1) + 2 \cdot \lambda = 0 \quad (1)$$

$$\frac{\partial L}{\partial a_2} = 2 \cdot (a_2 - 2) - 2 \cdot \lambda = 0 \quad (2)$$

$$\frac{\partial L}{\partial \lambda} = a_1 - a_2 = 0 \quad (3)$$




Add (1) and (2)

$$2 \cdot (a_1 - 1) + 2 \cdot (a_2 - 2) = 0$$

Insert (3)

$$2 \cdot (a_1 - 1) + 2 \cdot (a_1 - 2) = 0$$

**Solution:** $4 \cdot a_1 - 6 = 0 \rightarrow a_1 = a_2 = \frac{3}{2}$

# Exercise 1: Energy Conservation

- A $Z^0$ decays in its rest frame into two massless particles with the opposite momentum (back to back) and same energy
- If the energies $E_i$ of the particles are measured with an uncertainty $\sigma$, this can be formulated as a constrained linear least square problem.
- Assuming perfect angular resolution this problem has only two parameters ($E_i$)

Start the virtual machine, open a shell, and copy and extract the exercise material:

```
$> cp /statistics-school/ConstrainedFits/ConstrainedFits.tgz .
```

- Unpack the downloaded archive:

```
$> tar xvzf ConstrainedFits.tgz
```

- Change into the unpacked directory and run the setup script:

```
$> cd Constrained
$> . setup.sh
```

- Do exercise **1a** (Energy conservation) **1b** (Invariant mass contraint) following the instructions on the exercise sheet! (first paper and pen, then run the macro)

# Exercise 1: Summary

- Extremum of

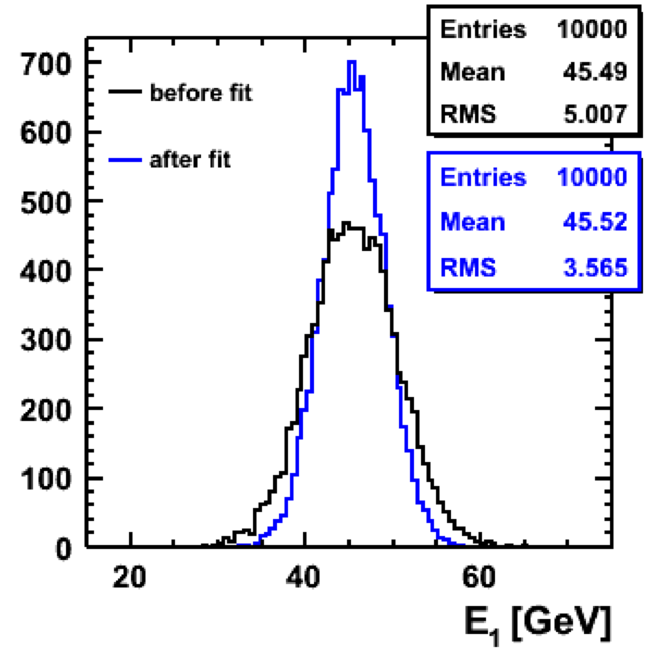
$$L(\vec{E}, \vec{a}) = \frac{(\hat{E}_1 - E_1)^2}{\sigma_1^2} + \frac{(\hat{E}_2 - E_2)^2}{\sigma_2^2} + 2 \cdot \lambda(E_1 + E_2 - M_Z)$$

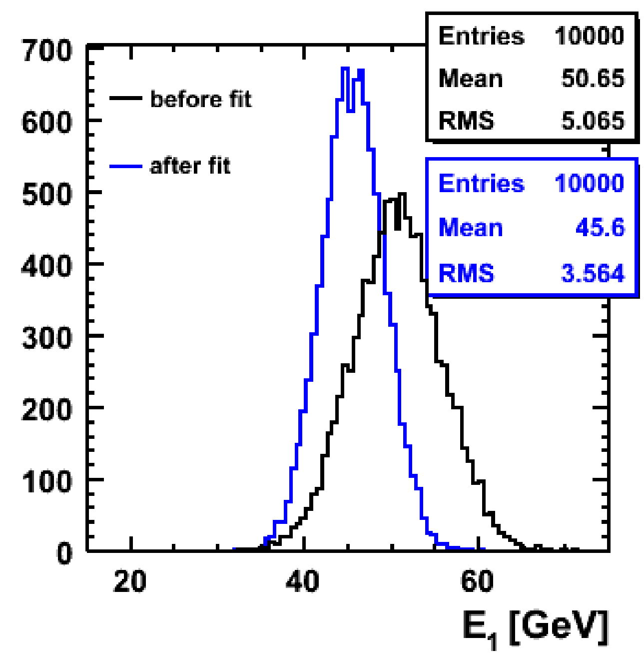
lies on diagonal in $E_1 - E_2$ plane defined by the constraint

- Solution is symmetric in $E_1$, $E_2$ if $\sigma_1 = \sigma_2$
- Energy resolution is improved as expected

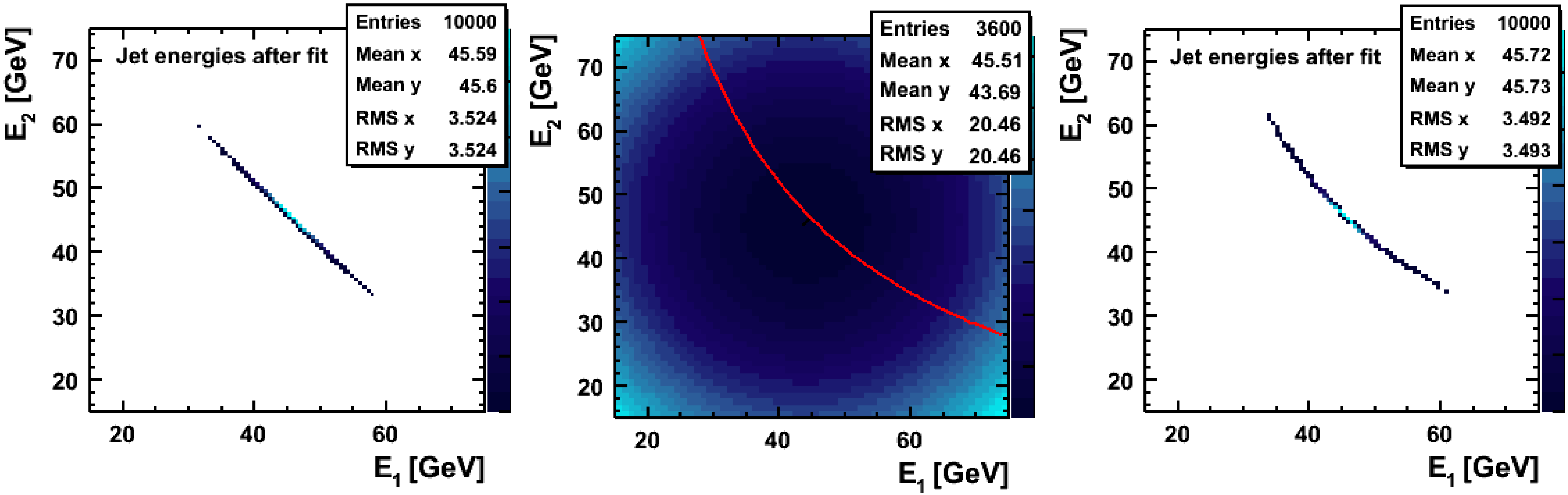
$$\sigma_{\text{fit}} = \sigma / \sqrt{2}$$

- Constraint fulfilled after fit
- Fit has **one** degree of freedom
- Perfectly gaussian input errors yield flat distribution of $\chi^2$-probability and perfect pulls (normalized corrections to $E_{1/2}$)
- Systematic shift of input energy is corrected by the fit



# Exercise 1: Summary




**Left:** Fitted jet energies after energy conservation constraint

**Middle:** $\chi^2$ profile for one event (blue) and invariant mass constraint (red line)

**Right:** Fitted jet energies after invariant mass constraint

→ Exercise 1a

→ Exercise 1b

# Non-linear Constraints

In our case: $\chi^2$ is quadratic in $a$ $\rightarrow$ derivatives are linear in $a$:

$$\chi^2 = (\vec{y} - \vec{a})^T V^{-1}[y](\vec{y} - \vec{a}) \rightarrow \frac{\partial \chi^2}{\partial \vec{a}} = -2 \cdot V^{-1}[y] \cdot (\vec{y} - \vec{a})$$

Constraint functions will in general be non-linear $\rightarrow$ make a Taylor expansion:

$$\vec{c}(\vec{a}) \approx \vec{c}(\vec{a^\star}) + \frac{\partial \vec{c}}{\partial \vec{a}} \cdot (\vec{a} - \vec{a^\star}) = \vec{c}(\vec{a^\star}) + A \cdot (\vec{a} - \vec{a^\star})$$

Fully linearized Lagrange function:

$$L \approx (\vec{y} - \vec{a})^T V^{-1}[y](\vec{y} - \vec{a}) + 2 \cdot \vec{\lambda}^T \cdot \vec{c}(\vec{a^\star}) + 2 \cdot \vec{\lambda}^T \cdot A \cdot (\vec{a} - \vec{a^\star})$$

Partial derivatives:

$$\frac{\partial L}{\partial \vec{a}} = -2 \cdot V^{-1}[y] \cdot (\vec{y} - \vec{a}) + 2 \cdot A^T \cdot \vec{\lambda} \stackrel{!}{=} 0$$

$$\frac{\partial L}{\partial \vec{\lambda}} = 2 \cdot \vec{c}(\vec{a^\star}) + 2 \cdot A \cdot (\vec{a} - \vec{a^\star}) \stackrel{!}{=} 0$$

In matrix form:

$$\begin{pmatrix} V^{-1}[y] \cdot \vec{y} \\ A \cdot \vec{a^\star} - \vec{c}(\vec{a^\star}) \end{pmatrix} = \begin{pmatrix} V^{-1}[y] & A^T \\ A & 0 \end{pmatrix} \cdot \begin{pmatrix} \vec{a} \\ \vec{\lambda} \end{pmatrix}$$

$\rightarrow$ Solve this equation to get a better solution, and iterate …

# Problems with Iterative Procedure

**The main problem:**

If local linear approximation of constraints is not very good, iterative procedure will make too small or too large steps

→ new solution may not be better (or even worse) than the old one

→ no convergence !

**Typical problems and ways to overcome/avoid them**:

- How to define if a step improves solution (constraints vs. $\chi^2$) → function of merit
- Step with unscaled length does not improve solution → step scaling
- Step rejected by "improvement criteria" (Maratos effect) → $2^{nd}$ order corrections

# Convergence

**How to qualify if a step is an improvement or not?**

**If current parameters away from hyperplane of fulfilled constraints** $\rightarrow$ ensure that each step reduces absolute values of constraints

**If current parameters are (almost) fulfilling the constraints** $\rightarrow$ ensure that objective function (e.g. $\chi^2$) is reduced

Construct new **function of merit**:

$$m_\mu(\vec{a}) = \chi^2(\vec{a}) + \mu \cdot \sum_{i=1}^{P} |c_i(\vec{a})|$$

with $\mu > 0$ (good estimate for $\mu$ is $\max(\lambda_i)$ )

One step is accepted if function of merit is reduced!

**Convergence:** two criteria have to be fulfilled

- Change of $\chi^2$ becomes small: $\Delta\chi^2 = \chi^2(\vec{a}) - \chi^2(\vec{a} + \Delta\vec{a}) < \epsilon_1$
- Absolute sum of constraints becomes small: $\sum_{i=1}^{P} |c_i(\vec{a})| < \epsilon_2$

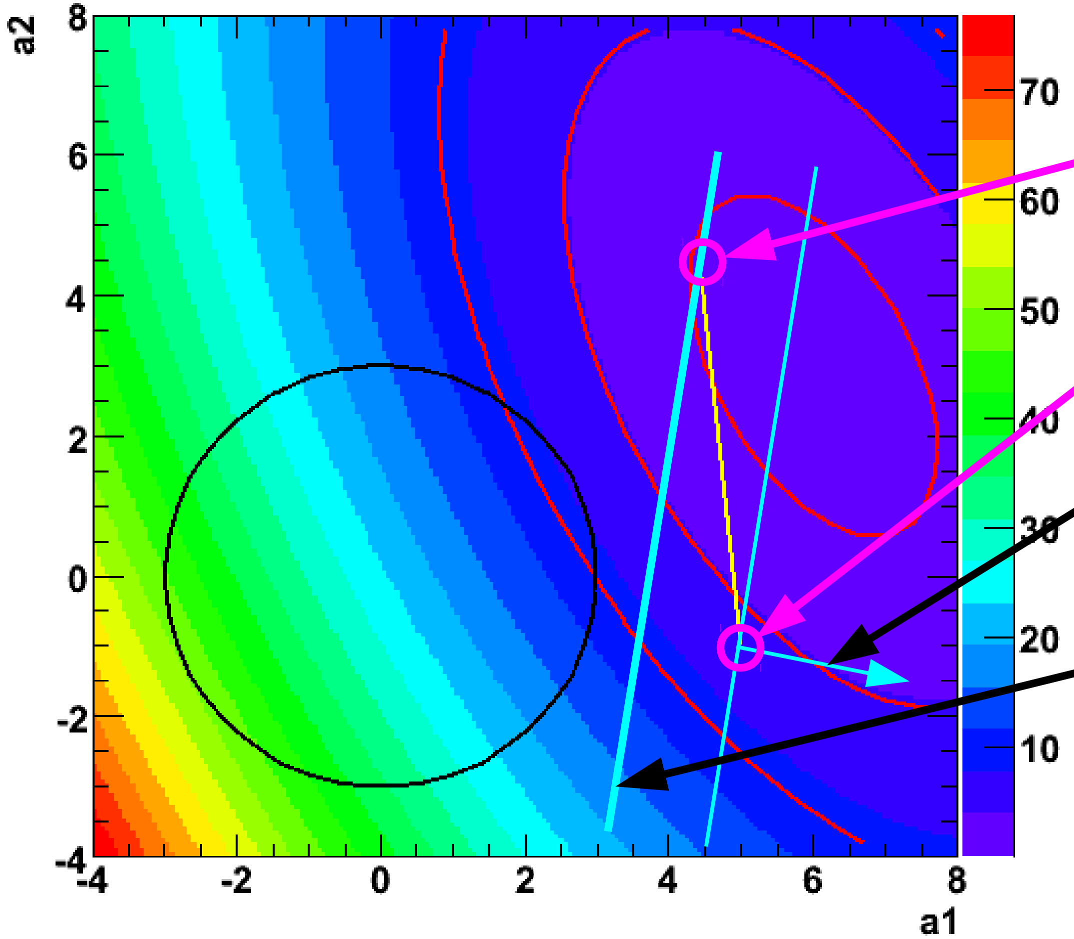
# Iterative Solution

**Example:** find minimum of

$$\chi^2 = (a_1 - 6, a_2 - 3) \begin{pmatrix} 3 & -0.5\sqrt{3\cdot 6} \\ -0.5\sqrt{3\cdot 6} & 6 \end{pmatrix}^{-1} \begin{pmatrix} a_1 - 6 \\ a_2 - 3 \end{pmatrix}$$

subject to

$$c(a_1, a_2) = a_1^2 + a_2^2 - 9 \stackrel{!}{=} 0$$




Solution of linearized problem local minima of $\chi^2$ on zero contour of linearized constraint

Starting parameters

Local gradient of constraint

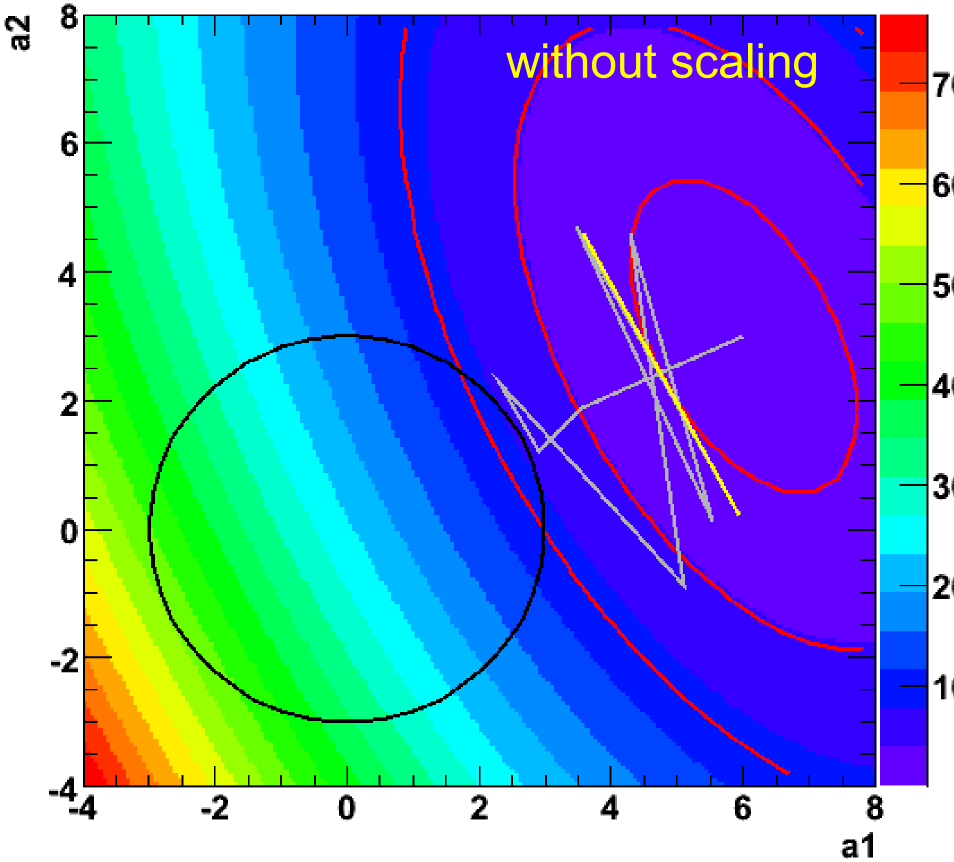
Zero contour of linearized constraint

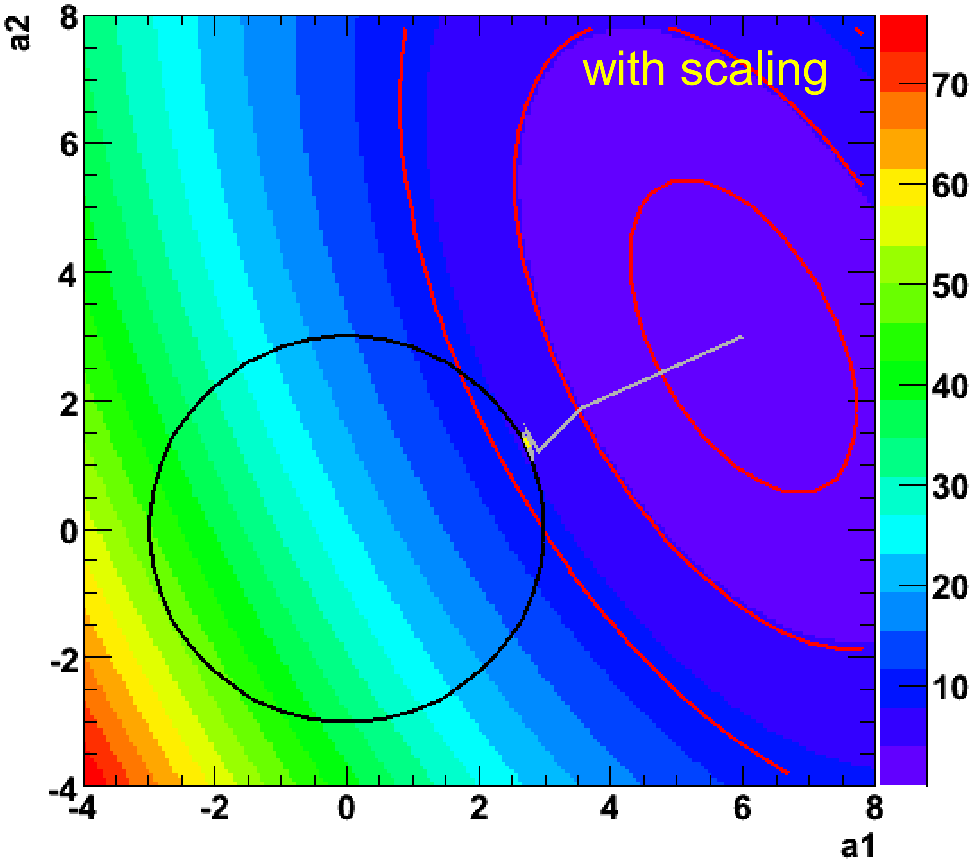
**Solution of problem:**
If step does not "improve" solution, make smaller step!

# Scaling of Step Length

For non-linear problems step after linearization does not improve solution (no reduction of function of merit)

→ **scale step length with $\alpha$** (e.g. $\alpha$ = 0.5)



# Remark: 2nd Order Corrections

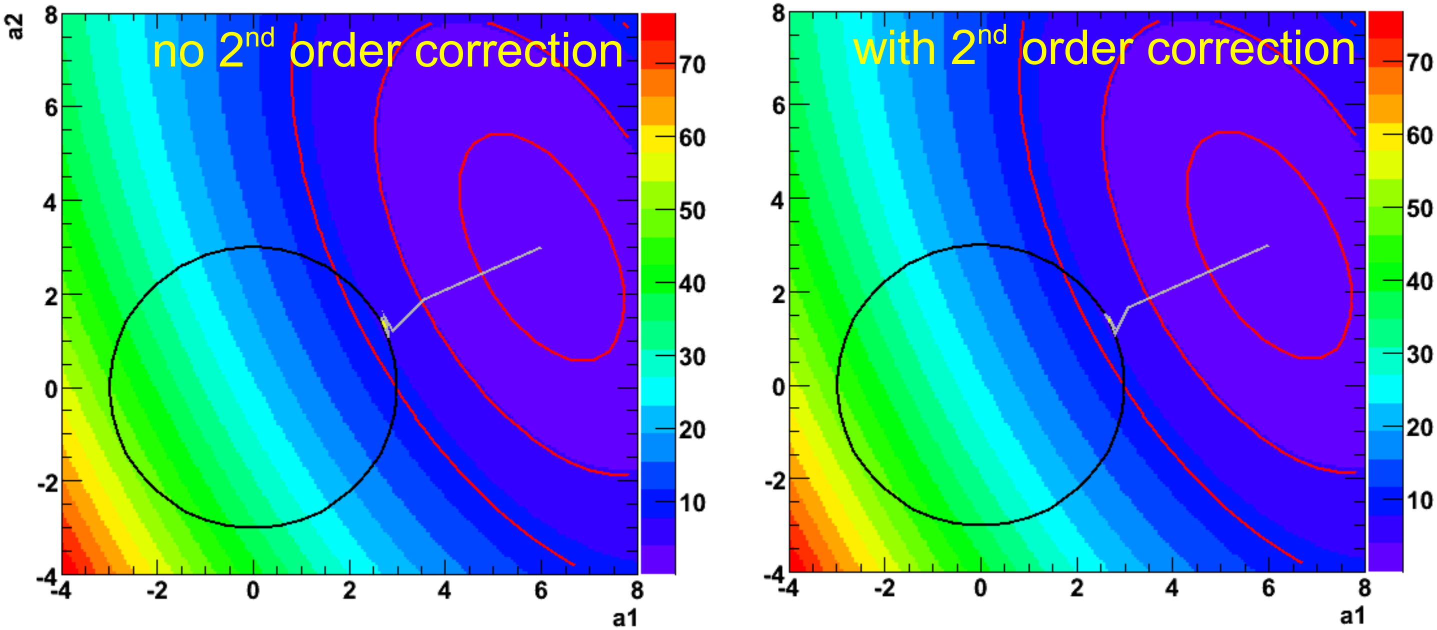
If constraints non-linear → one could suffer from Maratos effect, i.e. a step $\Delta\vec{a}_k$ calculated from the linearized problem might not improve the merit function → significant slow down of convergence

The following correction will reduce the merit function at least near the solution:

$$\Delta\vec{a'}_k = A_k^T (A_k A_k^T)^{-1} \vec{c}(\vec{a}_k + \Delta\vec{a}_k)$$

where $\Delta\vec{a}_k$ is the step, $\vec{c}(\vec{a})$ the constraints and $A$ the Jacobian of $\vec{c}$

This step is only tried once, since back tracking make no sense ( $\Delta\vec{a}_k$ is not steepest descent at $\vec{a}_k + \Delta\vec{a'}_k$ )




In this example "only" small improvement of convergence with 2nd order correction

# Application in High Energy Physics

**Possible applications of constrained fits:**

- Vertex constrained track fits
- Tracker Alignment
- Kinematic fits ...

**Kinematic Fits:**

- If measured jets are decay products of heavy particles → invariant mass of added 4-vectors should equal the mass of decaying particle
- Different parametrizations of final state momenta possible (see next slide)
- In general the measurements have to be shifted within their uncertainty to fulfill the constraint → one parameter per measurement

  **Problem can be formulated as constrained non-linear least square fit**

# Invariant Mass Constraints

Example for 2 jets from $W$ decay:

**Cartesian parametrization**:

$$m_W^2 - ((E_1 + E_2)^2 - (p_{x,1} + p_{x,2})^2 - (p_{y,1} + p_{y,2})^2 - (p_{z,1} + p_{z,2})^2) \stackrel{!}{=} 0$$

Advantage: simple calculation of derivatives, quadratic constraints

Disadvantage: measurements of different jet momentum components are correlated (some off-diagonal elements of the covariance matrix are non-zero)

**Different parametrization (here for massless jets):**

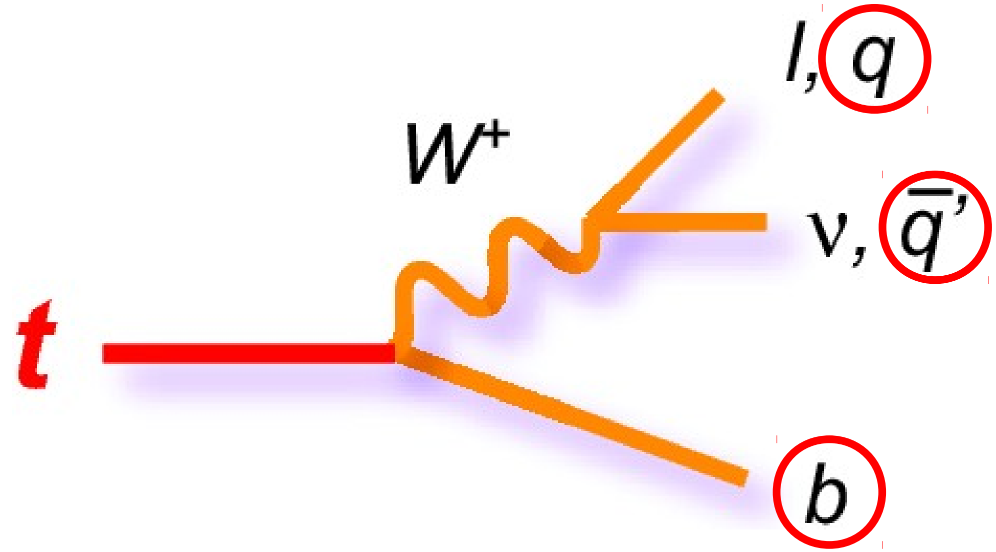
$$m_W^2 - \left( \left( \sum_{i=1}^{2} p_t \cosh \eta \right)^2 - \left( \sum_{i=1}^{2} p_t \cos \phi \right)^2 - \left( \sum_{i=1}^{2} p_t \sin \phi \right)^2 - \left( \sum_{i=1}^{2} p_t \sinh \eta \right)^2 \right) \stackrel{!}{=} 0$$

Advantage: measurements are independent (covariance matrix diagonal)

Disadvantage: highly non-linear constraints, derivatives more complicated

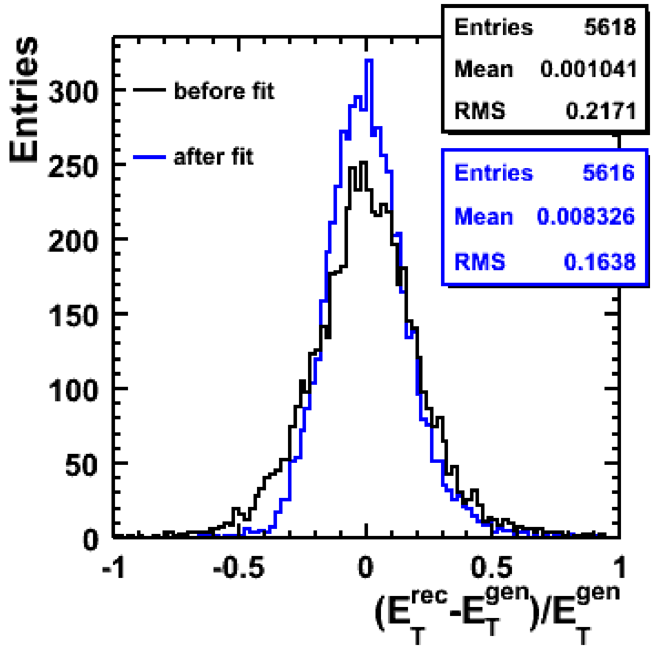
# Exercise 2: *W* and Top Mass Constraint

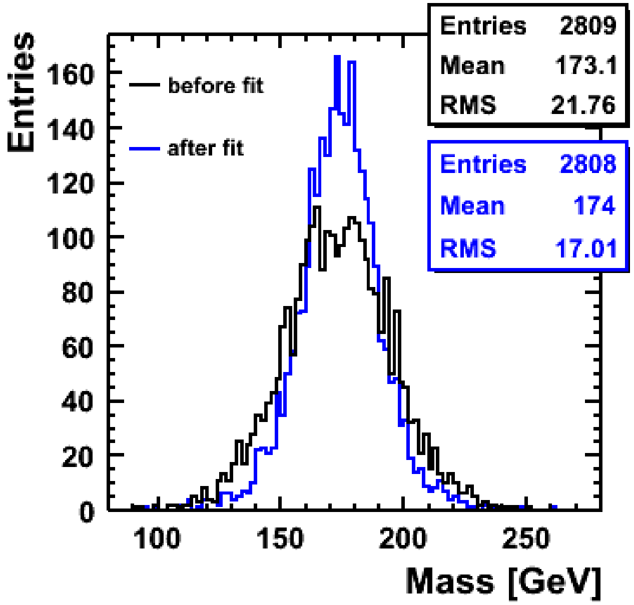
- Reconstruction of hadronic top decay
- Measured parameters: one *b* jet and two *W* jets = 9 parameters
- Mass constraints:
  - One *W* mass constrained → **2a**
  - Use known top mass in addition → **2b**



# Exercise 2: Summary

- Jet energy resolution is improved in the fit
- Top mass peak becomes narrower and top mass resolution is improved
- Jet $\eta$ and $\phi$ resolution not significantly changed as the assumed uncertainties are small
- Small peak at small $\chi^2$-probability: e.g. from events far from the assumed $W$ mass
- $\chi^2$ follows curve for **two** d.o.f in case of second constraint (2b)
- Pulls slightly distorted in such a more "realistic" scenario

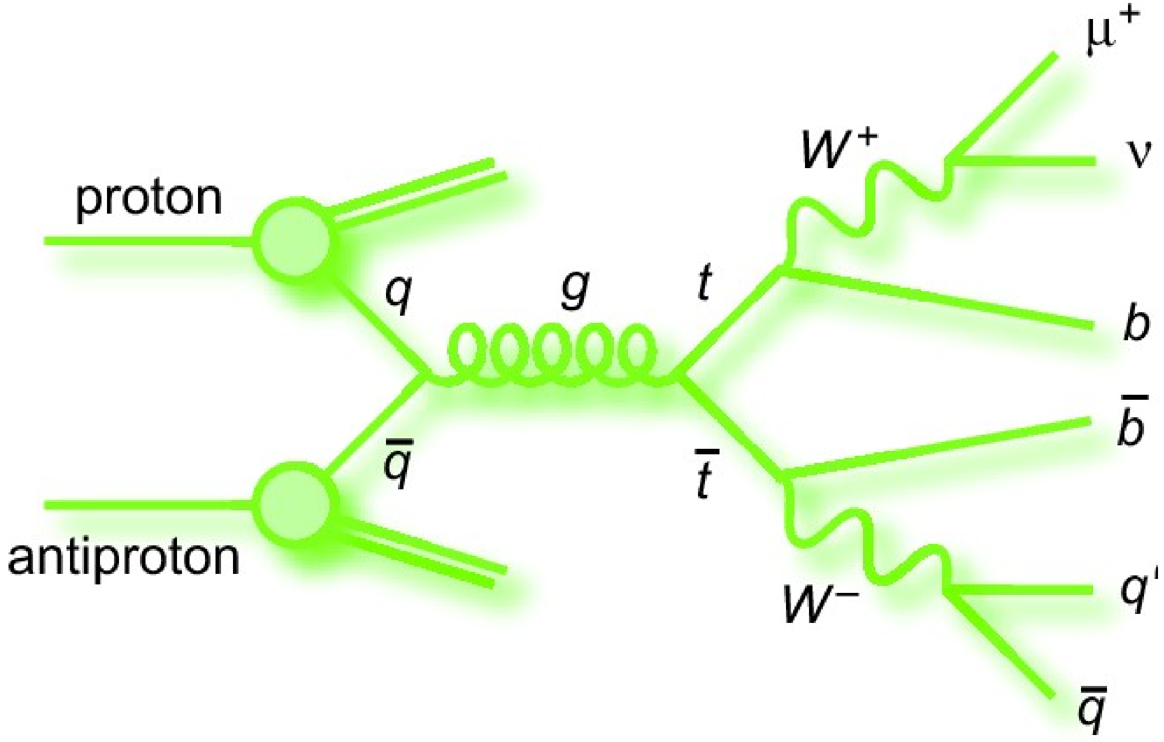


# Kinematic Fits of Whole Events

If a whole event is fitted the momentum balance can be used as **additional two constraints** (if the initial state has small transverse momentum)

In case of hard initial state radiation (ISR) the momentum balance of the hard process is broken → the ISR jets have to be taken into account




**Example:** semileptonic $t\bar{t}$ events at Tevatron

# Unmeasured Parameters

**General Problem:**

- $N$ measurements $\vec{y}$ and parameters $\vec{a}$ as before
- Unmeasurable particles (e.g. $\nu$) $\rightarrow$ $M$ additional unmeasured parameters $\vec{b}$
- $P$ constraints $\vec{c}(\vec{a}, \vec{b}) = 0$
- $P > M \rightarrow$ over-constrained problem

**Kinematic fit can be used to reconstruct unmeasured particle**

Linearized function $L$:

$$L \approx (\vec{y} - \vec{a})^T V^{-1}[y](\vec{y} - \vec{a}) + 2\vec{\lambda}^T(\vec{c}(\vec{a^\star}, \vec{b^\star}) + A \cdot (\vec{a} - \vec{a^\star}) + B \cdot (\vec{b} - \vec{b^\star}))$$

with Jacobian $A$ of constraints $\vec{c}$ with respect to **measured** parameters $\vec{a}$

and Jacobian $B$ of constraints $\vec{c}$ with respect to **unmeasured** parameters $\vec{b}$

and

$$\vec{c}(\vec{a}, \vec{b}) \approx \vec{c}(\vec{a^\star}, \vec{b^\star}) + A \cdot (\vec{a} - \vec{a^\star}) + B \cdot (\vec{b} - \vec{b^\star})$$

where all derivatives and function values are evaluated at starting values $\vec{a^\star}$ and $\vec{b^\star}$

# Unmeasured Parameters (cont.)

**Vanishing derivatives:** Overall $N + P + M$ coupled equations

$$\begin{pmatrix} V^{-1}[y] \cdot \vec{y} \\ 0 \\ A \cdot \vec{a^\star} + B \cdot \vec{b^\star} - \vec{c}(\vec{a^\star}, \vec{b^\star}) \end{pmatrix} = \begin{pmatrix} V^{-1}[y] & 0 & A^T \\ 0 & 0 & B^T \\ A & B & 0 \end{pmatrix} \cdot \begin{pmatrix} \vec{a} \\ \vec{b} \\ \vec{\lambda} \end{pmatrix}$$

with solution $\vec{a}$ and $\vec{b}$ which are new approximation of solution

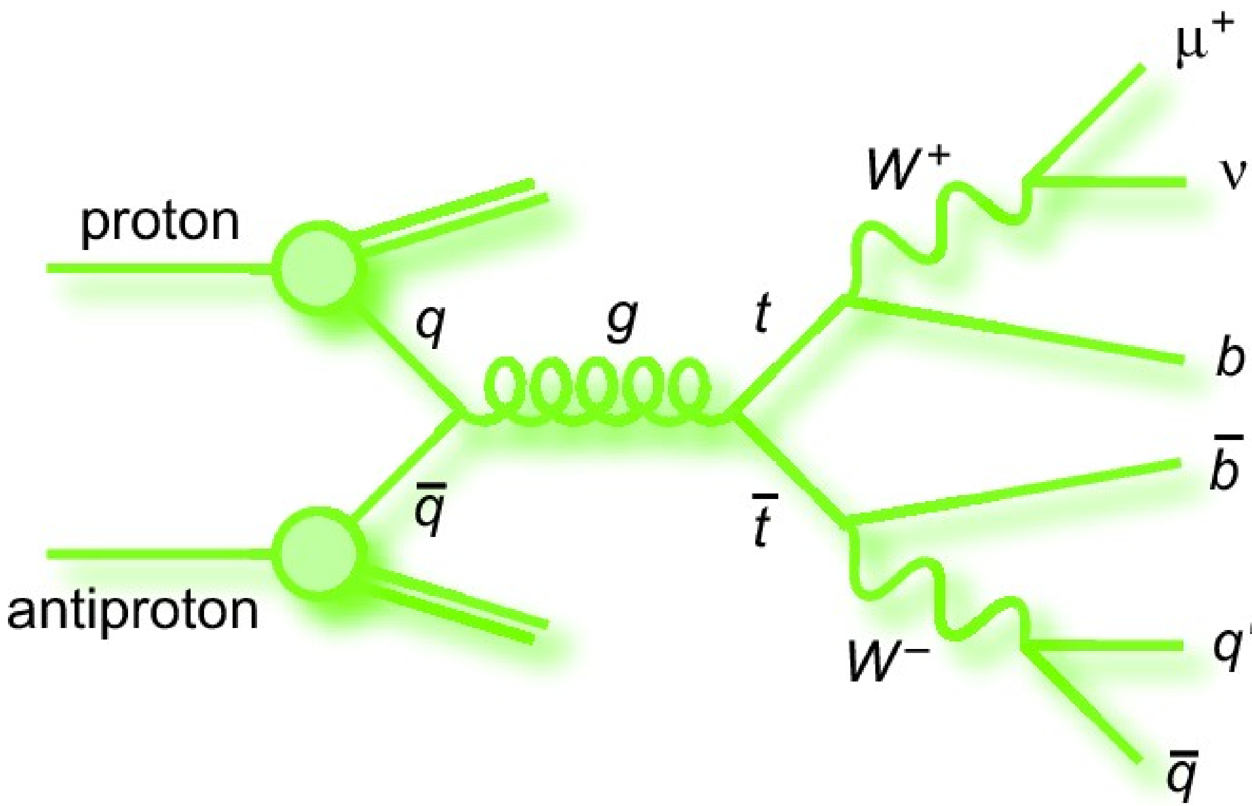
→ The matrix which has to be inverted has a special structure

→ Special algorithms make use of this structure which can save lots of computing resources (depending of the size of the problem)

# Exercise 3: Full Reconstruction of $t\bar{t}$

Reconstruction of hadronic and semileptonic $t\bar{t}$ events

- Both tops decay hadronically: Perform fit requiring an equal top mass on both branches → **3a**
- Semi-leptonic decay: Count parameters and find possible setup for an over-constrained fit → **3b**

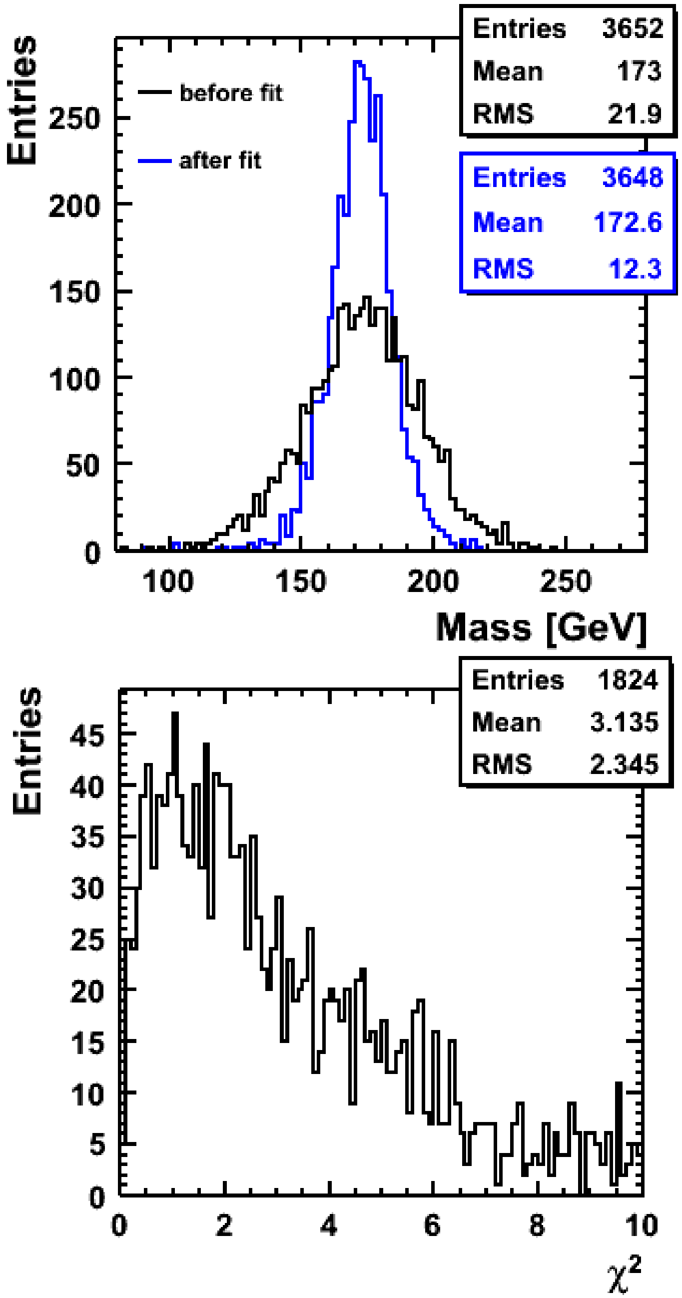


# Exercise 3a: Summary

- Both tops decaying hadronically:

  $2 \times W$ mass + equal top mass = **3 constr.**
- Same mass constraint brings gain in top mass width and resolution w.r.t. exercise 2a
- Energy resolution of jets profits from the fit as before
- $\chi^2$ as for **three** d.o.f. but some events accumulate at low fit probability

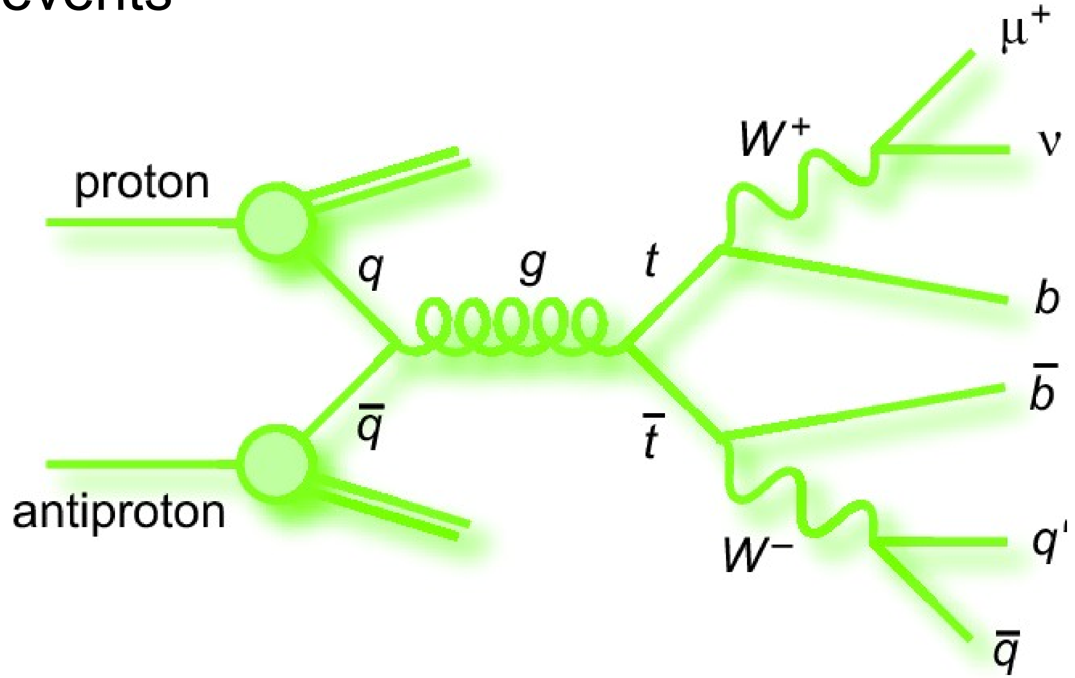
(**Remember:** No combinatorics included here!)



# Exercise 3b: Summary

- Reconstruction of semileptonic $t\bar{t}$ events
- Measured parameters:
  - Two $b$ jets
  - Two $W$ jets
  - One lepton

  $\Big\}$ 15 parameters
- Unmeasured parameters:
  - One neutrino = 3 parameters



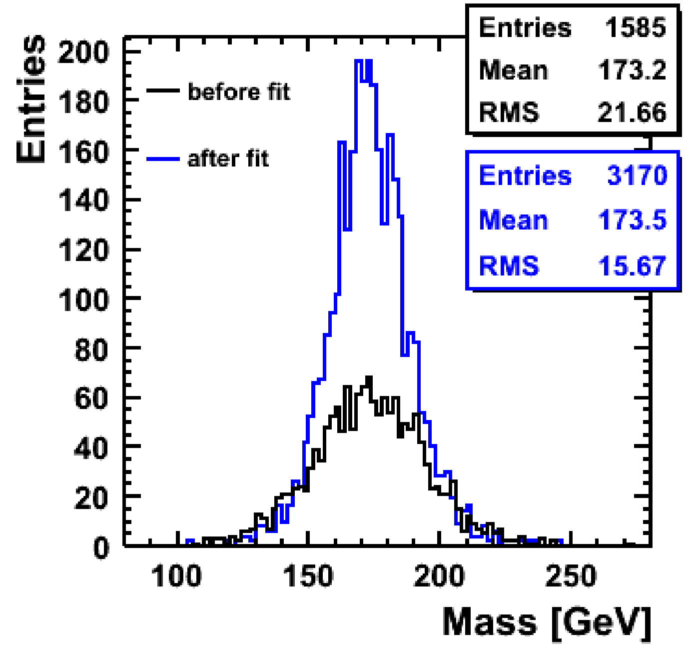

- Mass constraints:
  - Two $\times$ $W$ mass
  - Two $\times$ $p_T$ momentum balance
  - Zero, one (equality of two masses) or two $\times$ top mass

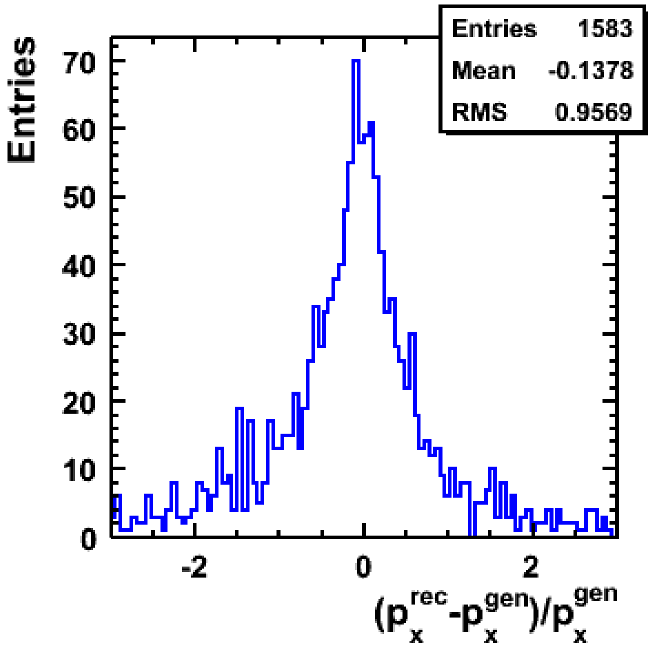
  $\Big\}$ 4 ... 6 constraints

**→ Need at least four constraints for kinematic fit!**

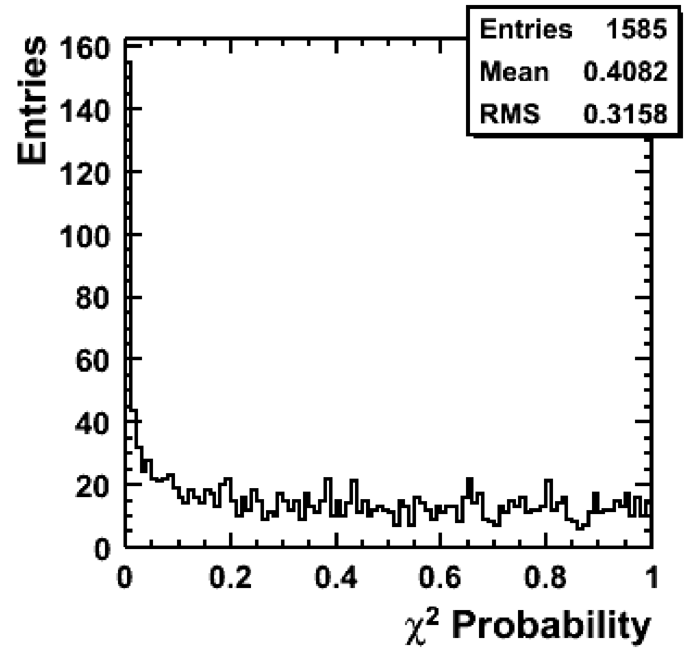
# Exercise 3b: Summary

- Use 2×$W$-mass + $p_T$-balance + equal top mass = 5 constraints
- For some events fit does not converge!
- Top mass reconstruction: good improvement by the fit
- Neutrino momentum ($p_x$, $p_y$, $p_z$) reconstructed with some width



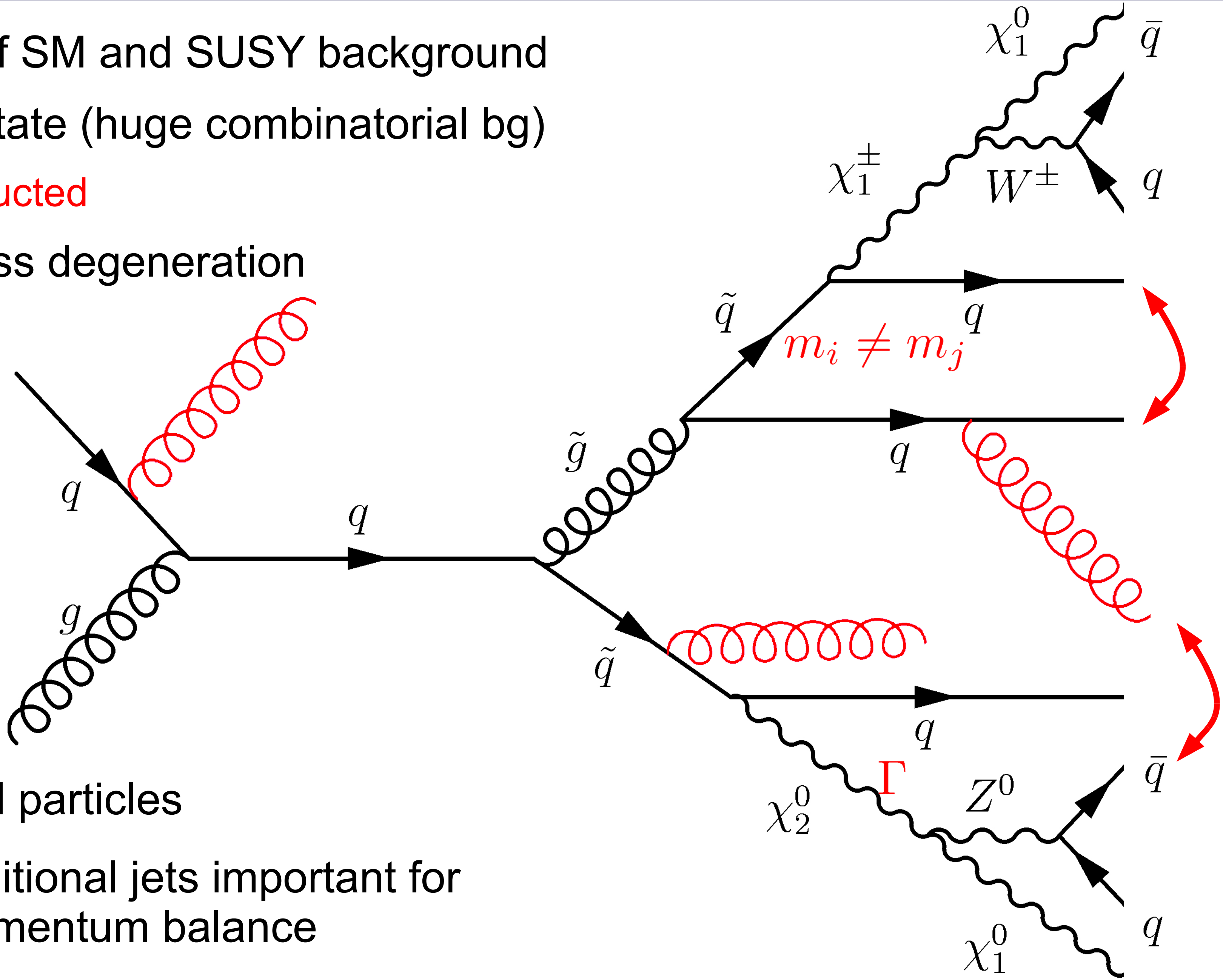




Fit probability for **two** d.o.f.: not ideally distributed, reflects complexity of the problem



# Example: SUSY Events

- Suppression of SM and SUSY background
- 7 jets in final state (huge combinatorial bg)
  - all reconstructed
- No perfect mass degeneration




- Width of virtual particles
- +FSR
- +ISR

additional jets important for momentum balance

# Jet Combinatorics

In case of many jets there are a lot of possible jet combinatorics

**Example:** *full hadronic $t\bar{t}$ events with 6 jets, two W jet pairs and two identical cascade branches*

Without information from *b*-tagging this leads to

$$\frac{6!}{2 \cdot 2} \cdot \frac{1}{2} = 90$$

combinations. With *b*-tagging:

$$2 \cdot \frac{4!}{2 \cdot 2} \cdot \frac{1}{2} = 6$$

**Another example:** *full hadronic SUSY event with 7 jets, two W jet pairs and different cascades* $\rightarrow$ 1260 combinations !

# Finite Width of Constraining Mass

If a mass of a constraint has a non-negligible width

→ Add a new parameter $x$ to the model

$x$ is scaling factor of constraining mass → new constraint has the form:

$$(x \cdot m)^2 - M_{\text{inv}}^2(\vec{a}, \vec{b}) \stackrel{!}{=} 0$$

The new parameter is treated as a new measurement per event with a variance according to the mass width:

$$\sigma_x^2 = \frac{\Gamma_m^2}{m^2}$$

→ A new $\chi^2$ term:

$$\chi_x^2 = \frac{(x-1)^2}{\sigma_x^2}$$

# Problems of Non-Linear Least Squares

**General problems:**

- Fit can converge at local (and not global) minimum
- Non linear problems can suffer from “Maratos effect”

**Alternative:** Formulation of constraints as additional $\chi^2$ term → “cost function”, e.g.

$$\left(\frac{M_{\text{inv}}^2(j_1, j_2, j_3) - M^2}{\sigma}\right)^2$$

with

$$\sigma = 2 \cdot M \cdot \Gamma_m$$

Minimize cost function: many possible algorithms (gradient, simplex, LBFGS, simulated annealing, genetic algorithm …)

**In general this quadrature is not advised**, but the procedure might be useful for very complex problems or for finding good starting values of the unmeasured fit parameters

# Remark: Inequality Constraints

Constraints can also be described by inequalities, e.g. if a parameter is restricted to positive values

**Two possibilities:**

- **Variable transformation:** mapping from finite to infinite parameter space, e.g. with trigonometric functions. Often this introduces more problems, e.g. additional saddle point or numeric uncertainties
- **Modification of Lagrangian multiplier method:** separation in active and non-active constraints → not simple

# Summary

- (Constrained) kinematic fits provide a powerful tool for event reconstruction. They can be used for:
    - Improvement of resolutions of measurements
    - Improvement of mass resolutions
    - Reconstruction of unknown parameters, like neutrino momenta
- Output is a $\chi^2$ which can be interpreted in terms of probabilities and can be used for event hypothesis classification
- Non linear problems have to be solved iteratively. The modification of the algorithm to achieve convergence is the hardest part!
- Minimization of scalar cost function might be useful to get good starting values of unmeasured parameter

# Literature

**Books:**

- V. Blobel and E. Lohrmann: ***Statistische und numerische Methoden der Datenanalyse*** (Teubner Studienbücher, 1998)
- J. Nocedal and S. J. Wright: ***Numerical Optimization***, 2nd Edition (Springer, 2006)
- L. Lyons: ***Statistics for nuclear and particle physicists*** (Cambridge Univ. Press 1986)
- …


Backup

# Remark: Scaling of Constraints

If objective function $\chi^2$ has to be minimized, subject to a number of constraints which are not fulfilled at starting parameters, it's possible that constraint values are of completely different magnitude!

Function of merit might be dominated by one single constraint

$$m_\mu(\vec{a},\vec{b}) = \chi^2(\vec{a},\vec{b}) + \mu \cdot \sum_{i=1}^{P} |c_i(\vec{a},\vec{b})| \approx \chi^2(\vec{a},\vec{b}) + \mu \cdot |c_k(\vec{a},\vec{b})|$$

For optimal performance of iterative algorithm → **scaling of constraint to same order of magnitude**

**Example:** invariant mass constraint in cascade decay

$$m(j_1 + j_2)^2 - m_1^2 = 0 \quad \text{and} \quad m(j_1 + j_2 + j_3)^2 - m_2^2 = 0$$

with $m_1 \ll m_2$

Squared mass difference of larger mass $m_2$ is dominant → normalize by expected uncertainty of constraint, e.g.

$$\Delta m_2^2 \approx 2 m_2$$