## Profiling Likelihood

$\mathcal{L}(D|\vec{\alpha},\vec{\beta})$ interested in $\vec{\alpha}$ not in $\vec{\beta}$

$\vec{\alpha}$ = nuisance parameter, eg BG contribution, efficiency

$\vec{\alpha}$ not perfectly known

$$\mathcal{L}(D|\vec{\alpha},\vec{\beta}) = \mathcal{L}(D|\vec{\alpha}\,\vec{\beta}) \times \mathcal{L}(\vec{\beta}\,|\,\vec{\beta})$$

a likelihood that supposedly comes from a subsidiary measurements — Not always realistic

Fit for both $\vec{\alpha}$ and $\vec{\beta}$

Profile likelihood $\mathcal{L}(D|\vec{\alpha},\hat{\hat{\vec{\beta}}})$ $\quad \hat{\hat{\vec{\beta}}}$ = value of $\vec{\beta}$ that maximizes $\mathcal{L}$ at a given $\vec{\alpha}$

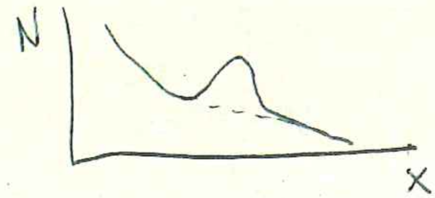
Use that as likelihood for $\vec{\alpha}$

Contrast with Bayesian approach where $\vec{\beta}$'s are integrated away —

## Extended max likelihood formalism



Fit a distribution ~~as~~ as sum of signal + background




S = number of signal events (to be fitted)

B = number of background events (to be fitted)

$P_S(x)$ = pdf for signal, normalized to $\underline{1}$

$P_B(x)$ = pdf for background, normalized to 1

$\vec{x} = \{x_1 \ldots x_N\}$ set of N values (events)

N = number of values (events)

$$p(N) = e^{-(S+B)} \frac{(S+B)^N}{N!} = \text{prob of having } N \text{ events}$$

$$P_i = \text{prob of event } i = \frac{S\,P_S(x_i) + B\,P_B(x_i)}{S+B}$$

Likelihood (extended!)

$$\mathcal{L} = p(N) \prod_{i=1}^{N} P_i =$$

unbinned

$$\mathcal{L} = e^{-(S+B)} \frac{(S+B)^N}{N!} \prod_{i=1}^{N} \left( \frac{S\,P_S(x_i) + B\,P_B(x_i)}{S+B} \right)$$

$$\mathcal{L} = e^{-(S+B)} \frac{(S+B)^N}{N!} \frac{\prod_{i=1}^{N} (S\,P_S(x_i) + B\,P_B(x_i))}{(S+B)^N}$$

ML8 ②

$$\mathcal{L} = e^{-(S+B)} \frac{1}{N!} \prod_{i=1}^{N} \left[ S P_S(x_i) + B P_B(x_i) \right] \quad (60)$$

$$-\log \mathcal{L} = S + B - \sum_{i=1}^{N} \log \left[ S P_S(x_i) + B P_B(x_i) \right] + \cancel{\log(N!)}$$

constant
ignore

$$\boxed{-\log \mathcal{L} = S + B - \sum_{i=1}^{N} \log \left[ S P_S(x_i) + B P_B(x_i) \right]}$$

# Constrained Fitting

Fit $N$ parameters $\vec{\alpha}$ with one (or more) constraint

For example $\alpha_2 + 3\alpha_1 = 4\alpha_3$ or $f(\vec{\alpha}) = \alpha_2 + 3\alpha_1 - 4\alpha_3 = 0$

In this case, easy! eg eliminate $\alpha_2 = 4\alpha_3 - 3\alpha_1$

then rewrite the likelihood (or $\chi^2$) as a function of $\{\alpha_1, \alpha_3, \alpha_4 \ldots\}$

For example, suppose we fit two tracks independently they corresponds to our "measurements" of 2 sets of trk parameters and cov matrices

Hypothesis: they come from a 2-body decay ie from the same space-point. Refit the two tracks using this additional information. Will improve the $\vec{p}$ resolution $\Rightarrow$ improve the invariant mass resolution. In this case the $f(\vec{\alpha}) = 0$ constraint is complicated, not easy to invert. Must do something else.

# CONSTRAINED FITS WITH LAGRANGE MULTIPLIERS

Recall fitting w/o constraint, with guess, Taylor expansion and iterations (if the problem is non-linear)

N- measurements $\vec{y}^m$, W = cov matrix

M- parameters $\vec{\alpha}$ ← This is what we want

Guess $\vec{\alpha}_0$ - $\vec{y}_0 = \vec{y}(\vec{\alpha}_0)$ $\delta\vec{y} = \vec{y}_m - \vec{y}_0$ $\delta\vec{\alpha} = \vec{\alpha} - \vec{\alpha}_0$

We had

$$\chi^2 = \underbrace{\underset{1\times N}{\delta\vec{y}^T}\ \underset{N\times N}{W^{-1}}\ \underset{N\times 1}{\delta\vec{y}}}_{1\times 1} - 2\underbrace{\underset{1\times N}{\delta\vec{y}^T}\ \underset{N\times N}{W^{-1}}\ \underset{N\times M}{A}\ \underset{M\times 1}{\delta\vec{\alpha}}}_{1\times 1} + \underbrace{\underset{1\times M}{\delta\vec{\alpha}^T}\ \underset{M\times N}{A^T}\ \underset{N\times N}{W^{-1}}\ \underset{N\times M}{A}\ \underset{M\times 1}{\delta\vec{\alpha}}}_{1\times 1}$$

with $A_{ij} = \frac{\partial y_i}{\partial \alpha_j}$ is N×M matrix

$\frac{\partial \chi^2}{\partial \alpha_j} = 0$ leads to $2A^T W^{-1} A\,\delta\vec{\alpha} - 2A^T W^{-1}\delta\vec{y} = 0$ (1)

Imagine that we have a set of p-constraints $\vec{f}(\vec{\alpha}) = 0$

Introduce a vector of p Lagrange multipliers

$\chi^2 = \chi^2_{OLD} + \vec{\lambda}^T \vec{f}(\vec{\alpha})$

Minimize now wrt <u>both</u> $\alpha$ and $\vec{\lambda}$

The terms that come from $\frac{\partial^2 \chi}{\partial \alpha_j}$ acquire an additional contribution $\lambda_r \frac{\partial f_r}{\partial \alpha_j}$ - New matrix $DF_{rj} = \frac{\partial f_r}{\partial \alpha_j}$ $\boxed{\text{P×M matrix}}$

We also get terms associated with $\frac{\partial \chi^2}{\partial \lambda_r} = 0$

Writing $\vec{\lambda}^T \vec{F}(\vec{\alpha}) = \lambda_r f_r(\vec{\alpha}) = \lambda_r f_r(\vec{\alpha}_0) + \lambda_r DF_{rj}\, \delta\alpha_j$

we get that $\dfrac{\partial \chi^2}{\partial \lambda_r} = 0$ reduces to

$$(DF)_{rj}\, \delta\alpha_j = -f_r(\vec{\alpha}_0)$$

All of these then can be rewritten as follows

$$\underbrace{\begin{bmatrix} \overset{M\times M}{[2A^T W^{-1} A]} & \overset{M\times P}{[DF^T]} \\ \underset{P\times M}{[DF]} & \underset{P\times P}{[0]} \end{bmatrix}}_{\substack{(M+P)\times(M+P) \\ \text{Call this matrix } B}} \underbrace{\begin{bmatrix} [\delta\alpha] \\ [\lambda] \end{bmatrix}}_{(M+P)\times 1} = \underbrace{\begin{bmatrix} \overset{M\times 1}{[2A^T W^{-1}\, \delta\vec{y}]} \\ \underset{P\times 1}{[-f(\vec{\alpha}_0)]} \end{bmatrix}}_{\substack{(M\times P)\times 1 \\ \text{Call this matrix } X}}$$

$$\boxed{\begin{bmatrix} \delta\vec{\alpha} \\ \lambda \end{bmatrix} = B^{-1} X}$$

Find the $\delta\vec{\alpha}$ and iterate.

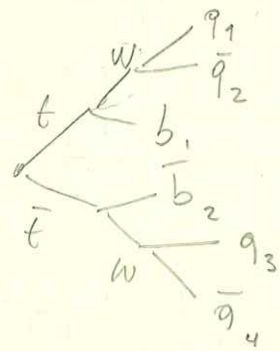
Note the covariance matrix is still $V = (A^T W^{-1} A)^{-1}$ as before

Note: this method does not work with most numerical minimizers (they are based on methods different from taking partial derivatives). They would happily minimize to $\lambda \to -\infty$. Another way of adding a constraint is by making it "soft" eg $\chi^2 = \chi^2 + \frac{f^2(\lambda)}{\sigma^2}$.

and pick a reasonably small $\sigma$

Often used when doing mass-constraint fits, and if the intermediate particle has a finite width this would actually be the right thing to do!! (modulo BW vs. Gaussian)

eg measure top mass




$q, b \longrightarrow$ jets

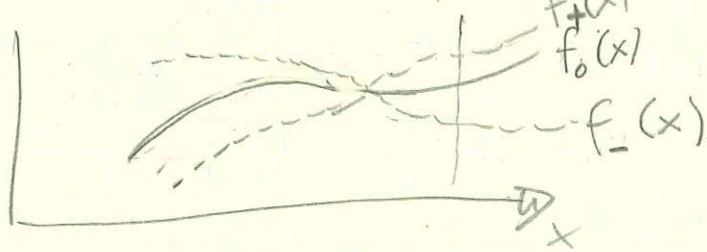
$$\chi^2 = \chi^2 + \frac{[M(j_3 j_4) - M_W]^2}{\sigma_W^2} + \frac{[M(j_1 j_2) - M_W]^2}{\sigma_W^2}$$

$$+ \frac{(M(j_1 j_2 b_1) - M(j_3 j_4 b_2))^2}{\sigma_t^2}$$

with $\sigma$'s of $\sim$ 1-2 GeV

## Morphing

Uncertainty in shape of PDF

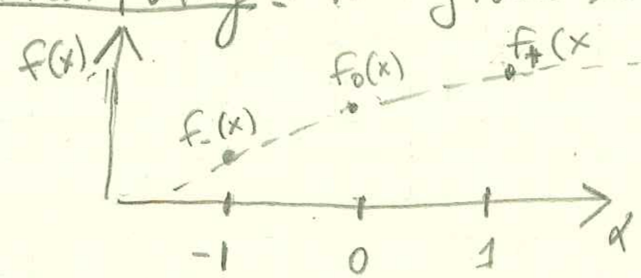



$f_0(x)$ = "central"

$f_\pm(x)$ = $\pm 1\,\sigma$ variation

Want to express $f(x)$ in terms of a nuisance parameter $\alpha$ so that interpolate smoothly btw $f_- \to f_0 \to f_+$

**Vertical Morphing** At given $x$




Several ways to interpolate _ Better if $\frac{df}{d\alpha}$ continous

eg $|\alpha|<1$, quadratic

$$f(x) = \frac{\alpha(\alpha-1)}{2} f_-(x) + 2\alpha f_0(x) + \frac{\alpha(\alpha+1)}{2} f_+(x)$$

$|\alpha|>1$ use some function, or (better) go linear with some slope

Then add Gaussian $\mu=0$ $\sigma=1$ for $\alpha$ in $\mathcal{L}$

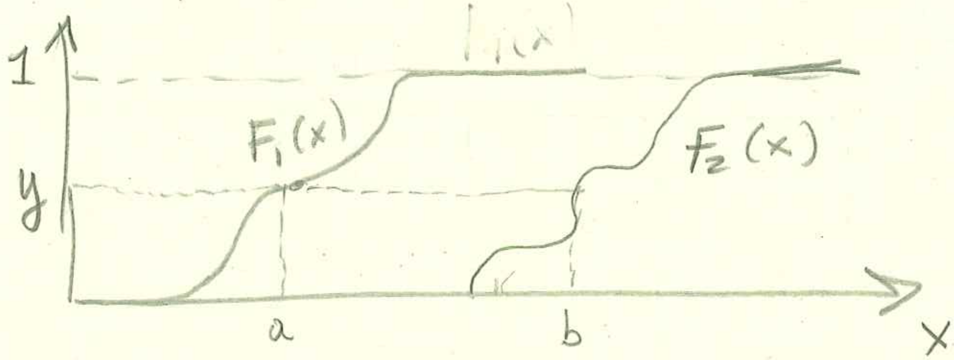
Generalizes to $>3$ $f(x)$'s, multidimension

# Horizontal Morphing

More robust, computationally messy
only 1D

eg interpolate btw $f_1(x)$ $f_2(x)$

$$F_1(x) = \int_{-\infty}^{x} f_1(t)\,dt \qquad F_2(x) = \int_{-\infty}^{x} f_2(t)\,dt$$




eg ½ way interpolation would be

$$F_{int}\left(x = \frac{a+b}{2}\right) = y$$

and $f_{int}(x)$ is such that

$$\int_{-\infty}^{x} f_{int}(t)\,dt = F_{int}(x)$$

Also other methods, eg morphing by moments. See linked papers