Practical Statistics for Physicists:
Learning to love the Covariance Matrix

Louis Lyons

Oxford &Imperial College

CMS expt at LHC

l.lyons@physics.ox.ac.uk

CERN School

Sept 2024

THE PARADOX

Histogram with 100 bins

Fit with 1 parameter

S_{min}: χ^2 with NDF = 99 (Expected $\chi^2 = 99 \pm 14$)

For our data, $S_{min}(p_0) = 90$

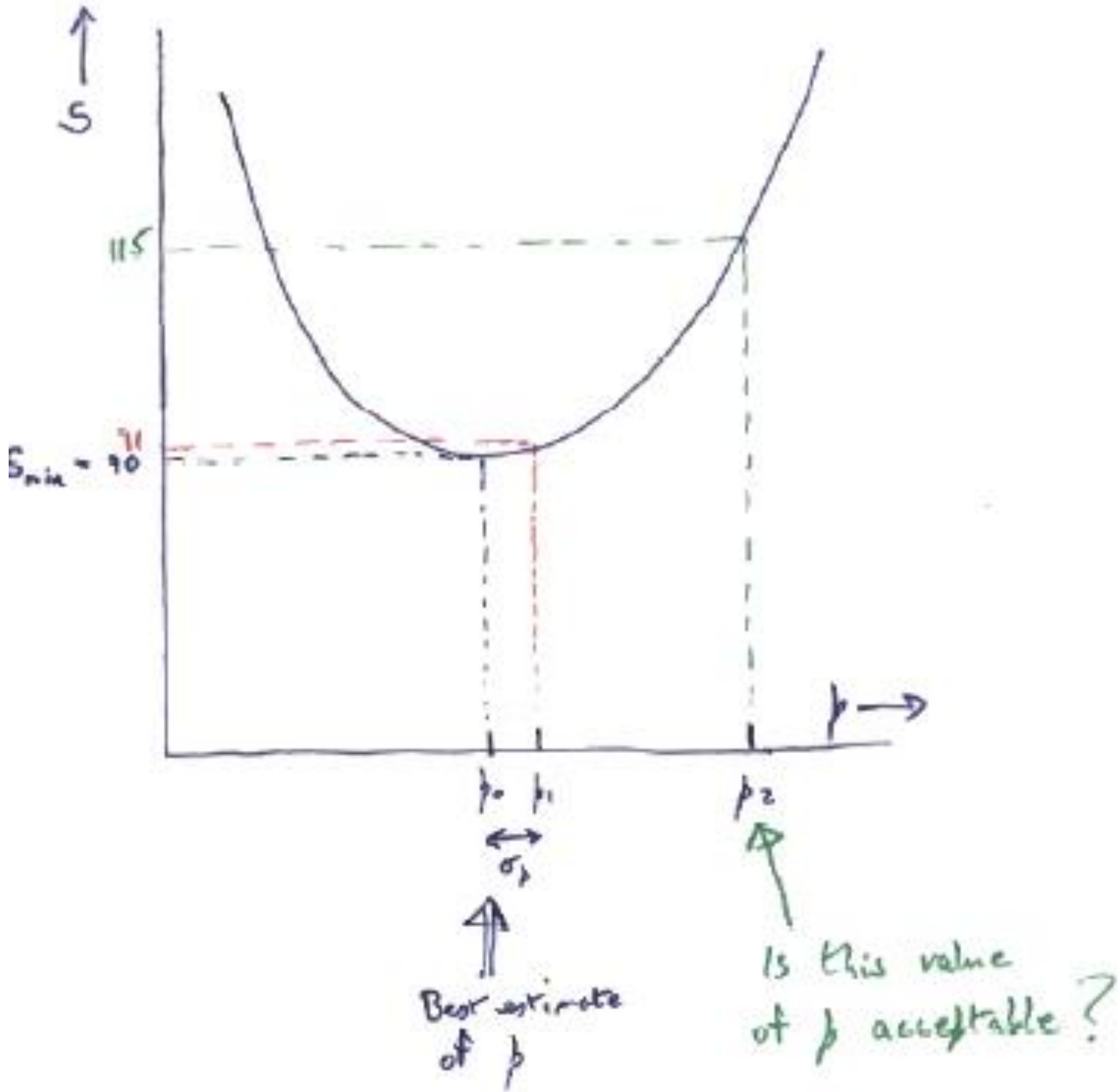
Is p_2 acceptable if $S(p_2) = 115$?

1) YES. Very acceptable χ^2 probability

2) NO. σ_p from $S(p_0 + \sigma_p) = S_{min} + 1 = 91$

But $S(p_2) - S(p_0) = 25$

So p_2 is 5σ away from best value


S
115
91
S_{min} = 90
p
p_0
p_1
p_2
σ_p
Best estimate of p
Is this value of p acceptable?

Correlations

Basic issue:

For 1 parameter, quote value and uncertainty

For 2 (or more) parameters,

(e.g. gradient and intercept of straight line fit)

quote values + uncertainties **+ correlations**

Just as the concept of variance for single variable is more general than Gaussian distribution, so correlation in more variables does not require multi-dim Gaussian

But more simple to introduce concept this way

Learning to love the Covariance Matrix

- Introduction via 2-D Gaussian
- Understanding covariance
- Using the covariance matrix

  Combining correlated measurements
- Estimating the covariance matrix

Reminder of 1-D Gaussian or Normal

14 *Experimental errors*

$$y = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\{-(x-\mu)^2/(2\sigma^2)\}$$

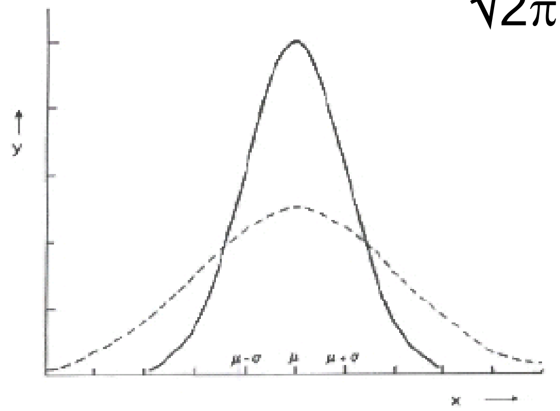
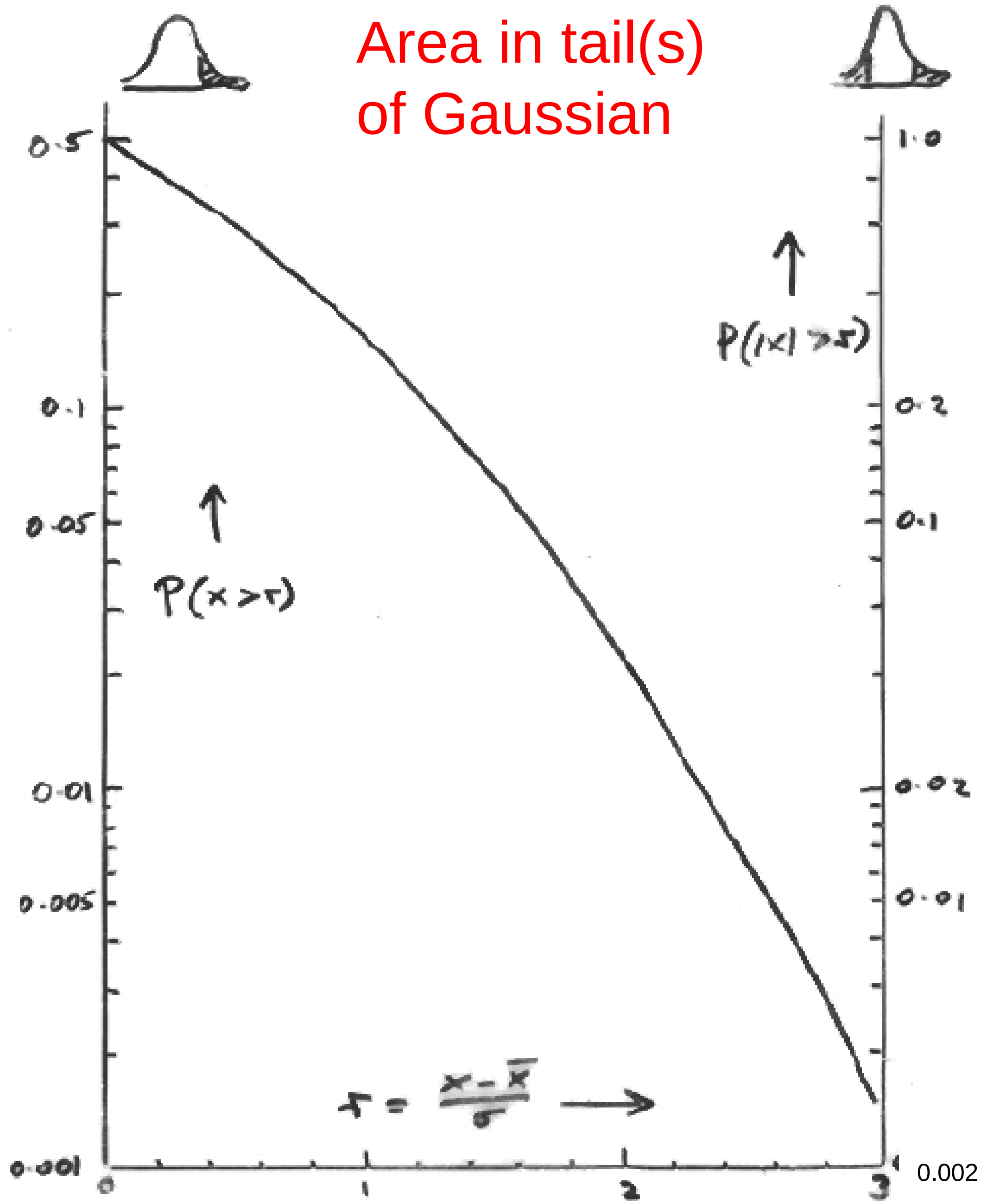


Fig. 1.5. The solid curve is the Gaussian distribution of eqn (1.14). The distribution peaks at the mean μ, and its width is characterised by the parameter σ. The dashed curve is another Gaussian distribution with the same values of μ, but with σ twice as large as the solid curve. Because the normalisation condition (1.15) ensures that the area under the curves is the same, the height of the dashed curve is only half that of the solid curve at their maxima. The scale on the x-axis refers to the solid curve.

Significance of σ

i) RMS of Gaussian = σ
(hence factor of 2 in definition of Gaussian)
ii) At $x = \mu \pm \sigma$, $y = y_{max}/\sqrt{e}$ ~0.606 y_{max}
(i.e. σ = half-width at 'half'-height)
iii) Fractional area within $\mu \pm \sigma$ = 68%
iv) Height at max = $1/(\sigma\sqrt{2\pi})$

Area in tail(s) of Gaussian



Gaussian in 2-variables

$$P(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_x} e^{-\frac{1}{2}\frac{x^2}{\sigma_x^2}}$$

$$P(y) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_y} e^{-\frac{1}{2}y^2/\sigma_y^2}$$

$x + y$ uncorrelated $\Rightarrow$

$$P(x,y) = \frac{1}{2\pi} \frac{1}{\sigma_x \sigma_y} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)}$$

Down on $P(0,0)$ by $e^{-\frac{1}{2}}$ when

$$\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} = 1$$

Rewrite as

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} \frac{1}{\sigma_x^2} & 0 \\ 0 & \frac{1}{\sigma_y^2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 1$$

Invert $\Rightarrow$ ERROR MATRIX

$$\begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix}$$

Element E_{ij} - $<(x_i - \overline{x}_i)(x_j - \overline{x}_j)>$

Diagonal E_{ij} = variances

Off-diagonal E_{ij} = covariances

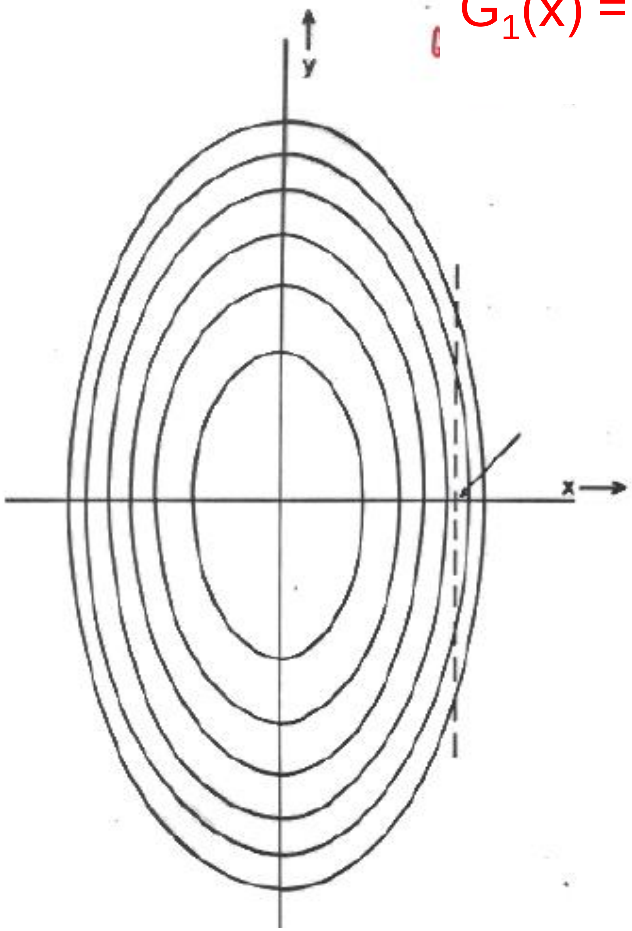
Towards the Covariance Matrix

x and y uncorrelated Gaussians

$P(x,y) = G_1(x)\, G_2(y)$

$G_1(x) = 1/(\sqrt{2\pi}\sigma_x)\, \exp\{-x^2/2\sigma_x^2\}$

$G_1(x) = 1/(\sqrt{2\pi}\sigma_y)\, \exp\{-y^2/2\sigma_y^2\}$



$P(x,y) = 1/(2\pi\sigma_x\sigma_y)\, \exp\{-0.5(x^2/\sigma_x^2 + y^2/\sigma_y^2)\}$

Specific example

$$\sigma_x = \frac{\sqrt{2}}{4} = .354 \qquad \sigma_y = \frac{\sqrt{2}}{2} = .707$$

Then factor of $e^{-\frac{1}{2}}$ when

$$8x^2 + 2y^2 = 1$$

Now introduce CORRELATIONS by 30° rotn

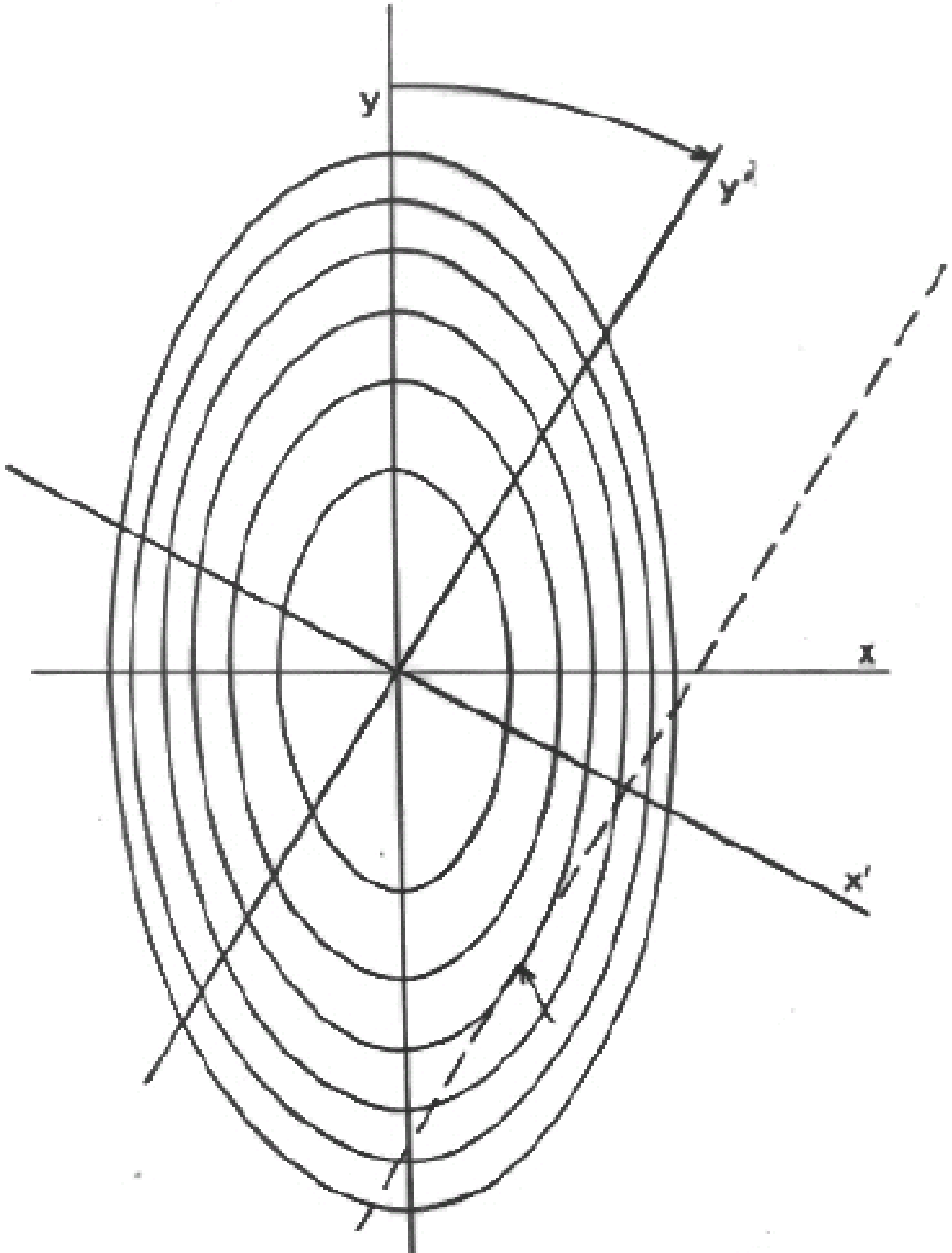
$$\frac{1}{2}\left[13x'^2 + 6\sqrt{3}\,x'y' + 7y'^2\right] = 1$$

$$\begin{pmatrix} \frac{13}{2} & \frac{3\sqrt{3}}{2} \\ \frac{3\sqrt{3}}{2} & \frac{7}{2} \end{pmatrix}$$

Inverse Covariance Matrix

$$\frac{1}{32} \times \begin{pmatrix} 7 & -3\sqrt{3} \\ -3\sqrt{3} & 13 \end{pmatrix}$$

Covariance Matrix


y
y'
x
x'

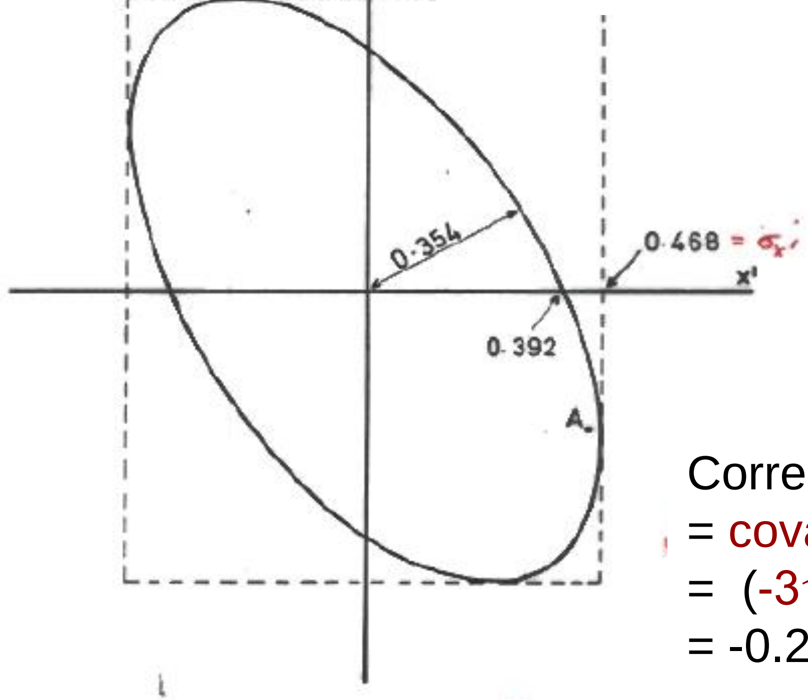
$8x^2 + 2y^2 = 1$

$0.5(13x'^2 + 6\sqrt{3}x'y' + 7y'^2) = 1$

$$\begin{pmatrix} 13/2 & 3\sqrt{3}/2 \\ 3\sqrt{3}/2 & 7/2 \end{pmatrix} \qquad (1/32)^* \begin{pmatrix} 7 & -3\sqrt{3}/2 \\ -3\sqrt{3}/2 & 13 \end{pmatrix}$$

Inverse covariance matrix

Covariance matrix



Correlation coefficient ρ

= covariance/σ(x')σ(y')

= $(-3\sqrt{3}/2)/\mathrm{sqrt}(7*13)$

= -0.27

$7/32 = (0.468)^2 = \sigma(x`)^2$

$1/6.5 = (0.392)^2$

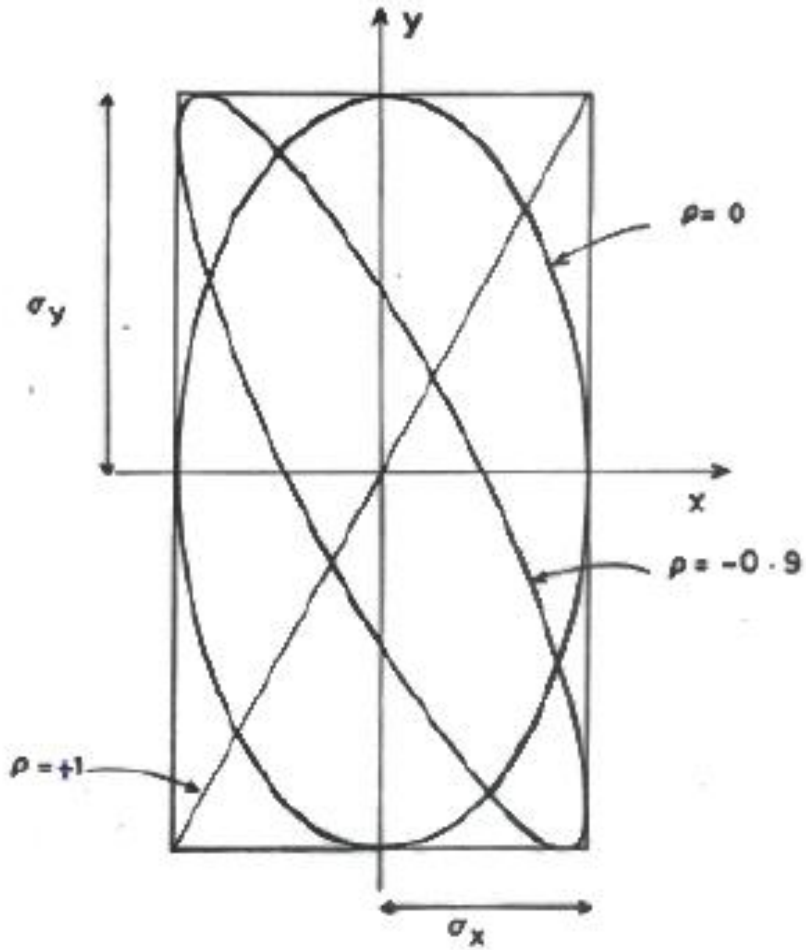
$1/8$ = eigenvalue of covariance matrix = $\sigma(x)^2$

$\left.\begin{matrix}\sigma_x \\ \sigma_y\end{matrix}\right\}$ constant

ρ varying

Covariance $\begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$

Covariance matrix,
ρ in range -1→+1



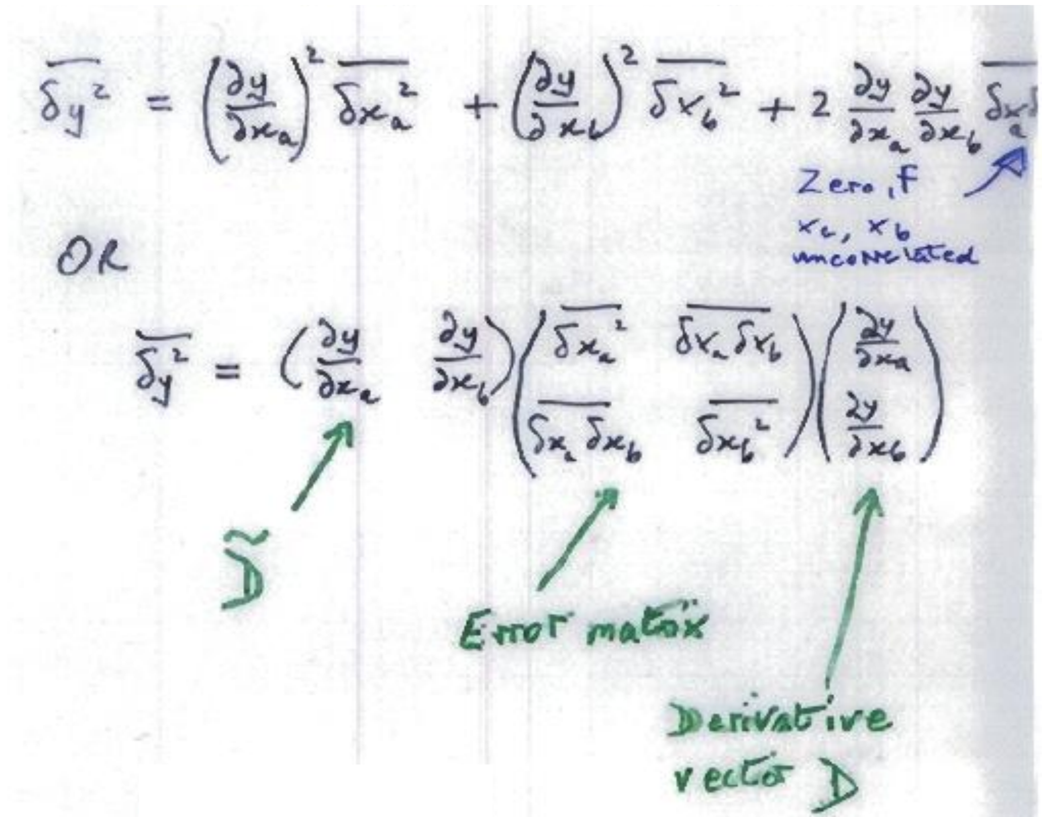
Using the Covariance Matrix

(i) Function of variables

$y = y(x_a, x_b)$

Given covariance matrix for x_a, x_b, what is σ_y ?

Differentiate, square, average



$\sigma_y^2 = \tilde{D}ED$

(ii) Change of variables $x_a = x_a(p_i, p_j)$

$x_b = x_b(p_i, p_j)$

e.g Cartesian to polars; or

Points in x,y → intercept and gradient of line

Given cov matrix for p_i, p_j, what is cov matrix for x_a, x_b ?

Differentiate, calculate $\delta x_a \delta x_b$, and average

$$\delta x_a = \frac{\partial x_a}{\partial p_i}\delta p_i + \frac{\partial x_a}{\partial p_j}\delta p_j \qquad (+ \text{ sim for } x_b)$$

Then

$$\overline{\delta x_a^2} = \left(\frac{\partial x_a}{\partial p_i}\right)^2 \overline{\delta p_i^2} + \left(\frac{\partial x_a}{\partial p_j}\right)^2 \overline{\delta p_j^2} + 2\frac{\partial x_a}{\partial p_i}\frac{\partial x_a}{\partial p_j}\overline{\delta p_i \delta p_j}$$

$$\overline{\delta x_a \delta x_b} = \frac{\partial x_a}{\partial p_i}\frac{\partial x_b}{\partial p_i}\overline{\delta p_i^2} + \frac{\partial x_a}{\partial p_j}\frac{\partial x_b}{\partial p_j}\overline{\delta p_j^2} + \left(\frac{\partial x_a}{\partial p_i}\frac{\partial x_b}{\partial p_j} + \frac{\partial x_a}{\partial p_j}\frac{\partial x_b}{\partial p_i}\right) \times \overline{\delta p_i \delta p_j}$$

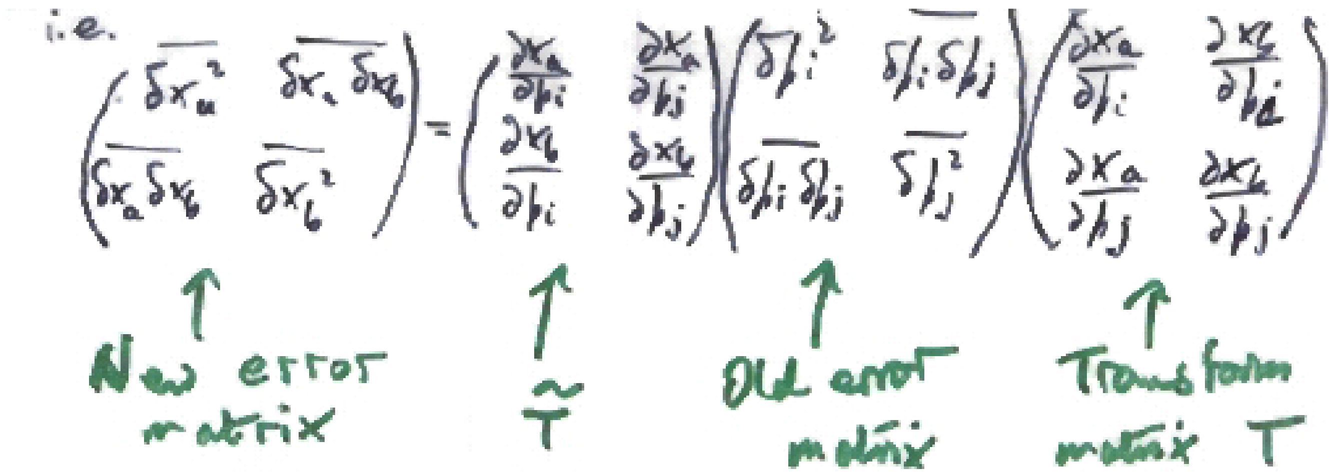
$+ \overline{\delta x_b^2}$ like $\overline{\delta x_a^2}$

N.B. Change of variables does not have to be $N \to N$

e.g. straight line fit involves $N \to 2$

Then i) & ii) are both examples of $N \to M$ ($M \leq N$)

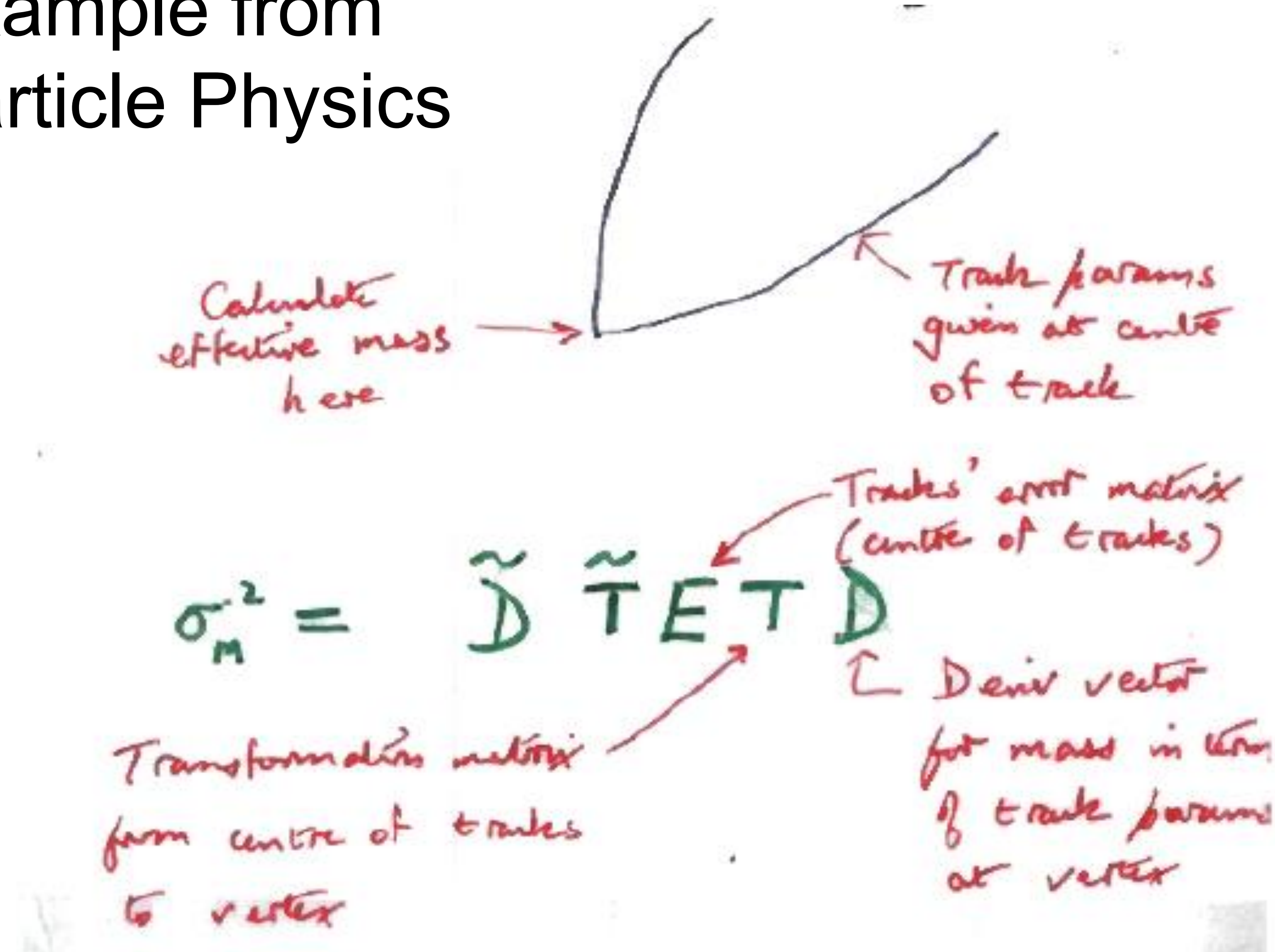
where $M = 1$ in i) $M = N$ in ii)



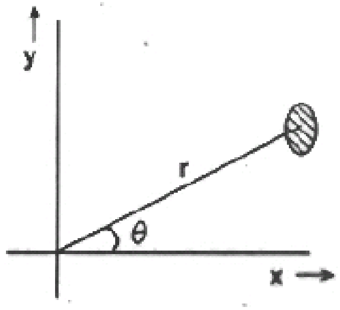
$$E_x = \tilde{T} E_p T$$

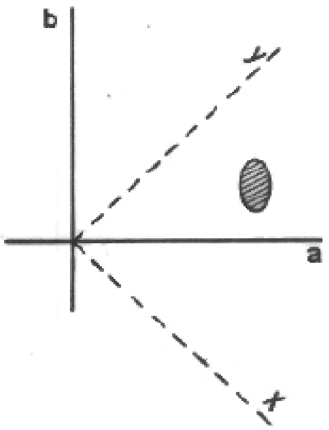
BEWARE!

Example from Particle Physics



Examples of correlated variables





Using the Covariance Matrix

COMBINING RESULTS

If $a_i \pm \sigma_i$ are independent:

Minimise $S = \sum \left(\frac{a_i - \hat{a}}{\sigma_i}\right)^2$

$$\Rightarrow \quad \hat{a} = \frac{\sum a_i w_i}{\sum w_i} \qquad w_i = 1/\sigma_i^2$$

Now $a_i \pm \sigma_i$ are correlated with error matrix $\underline{\underline{E}}$

$$\underline{\underline{E}} = \begin{pmatrix} \sigma_1^2 & cov(1,2) & cov(1,3) & \cdots \\ cov(1,2) & \sigma_2^2 & cov(2,3) & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

$$S = \sum_{i,j} (a_i - \hat{a}) \underline{\underline{E}}^{-1}_{ij} (a_j - \hat{a})$$

↑ INVERSE ERROR MATRIX

N.B. $\hat{a}$ CAN LIE OUTSIDE a_i

$\sigma_a \rightarrow 0$ AS $\rho \rightarrow \pm 1$

$$\underline{\underline{E}}^{-1} = \begin{pmatrix} 1/\sigma_1^2 & 0 & 0 & \cdots \\ 0 & 1/\sigma_2^2 & 0 & \\ \vdots & \vdots & \vdots & \end{pmatrix}$$ FOR UNCORRELATED

BLUE

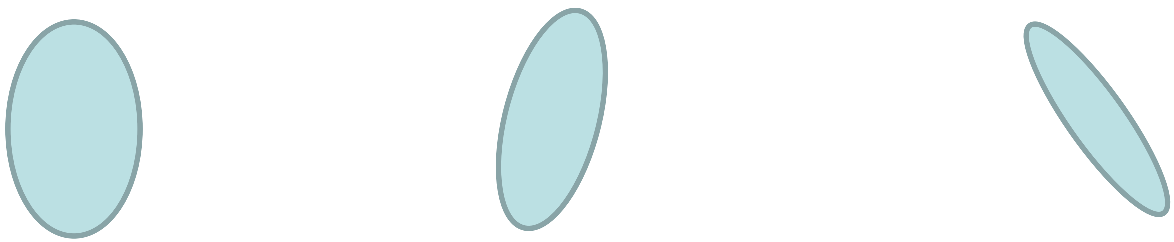
Best **L**inear **U**nbiassed **E**stimate

Combine several possibly correlated estimates of same quantity

e.g. v_1, v_2, v_3

Covariance matrix

$$\begin{pmatrix} \sigma_1^2 & cov_{12} & cov_{13} \\ cov_{12} & \sigma_2^2 & cov_{23} \\ cov_{13} & cov_{23} & \sigma_3^2 \end{pmatrix}$$



Uncorrelated   Positive correlation   Negative correlation

$$cov_{ij} = \rho_{ij}\, \sigma_i\, \sigma_j \quad \text{with} \quad -1 \le \rho \le 1$$

Lyons, Gibault + Clifford
NIM A270 (1988) 42

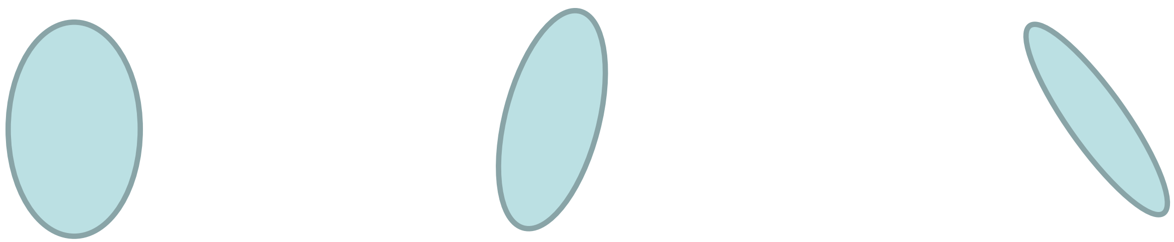
BLUE

Best Linear Unbiassed Estimate

Combine several possibly correlated estimates of same quantity

e.g. v_1, v_2, v_3

Covariance matrix
$$\begin{pmatrix} \sigma_1^2 & cov_{12} & cov_{13} \\ cov_{12} & \sigma_2^2 & cov_{23} \\ cov_{13} & cov_{23} & \sigma_3^2 \end{pmatrix}$$



Uncorrelated   Positive correlation   Negative correlation

$cov_{ij} = \rho_{ij}\,\sigma_i\,\sigma_j$ with $-1 \le \rho \le 1$

Lyons, Gibault + Clifford
NIM A270 (1988) 42

$v_{best} = w_1 v_1 + w_2 v_2 + w_3 v_3$ **L**inear

with $w_1 + w_2 + w_3 = 1$ **U**nbiassed

to give $\sigma_{best} = \min$ (wrt w_1, w_2, w_3) **B**est

For uncorrelated case, $w_i \sim 1/\sigma_i^2$

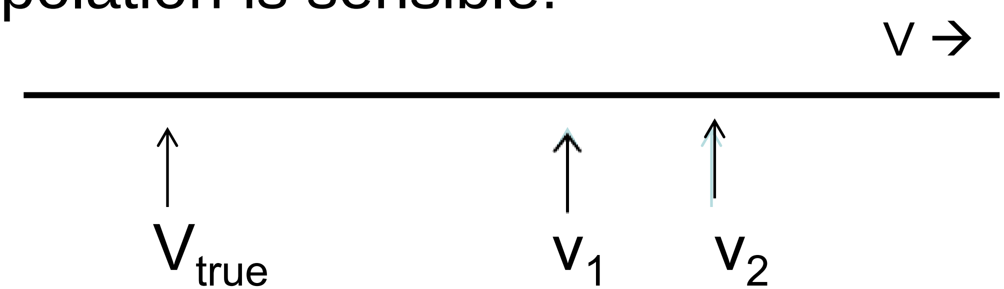
For correlated pair of measurements with $\sigma_1 < \sigma_2$

$v_{best} = \alpha v_1 + \beta v_2$ $\beta = 1 - \alpha$

$\beta = 0$ for $\rho = \sigma_1/\sigma_2$ (Smaller β → weights both >0)

$\beta < 0$ for $\rho > \sigma_1/\sigma_2$ i.e. extrapolation! e.g. $v_{best} = 2v_1 - v_2$

Extrapolation is sensible:



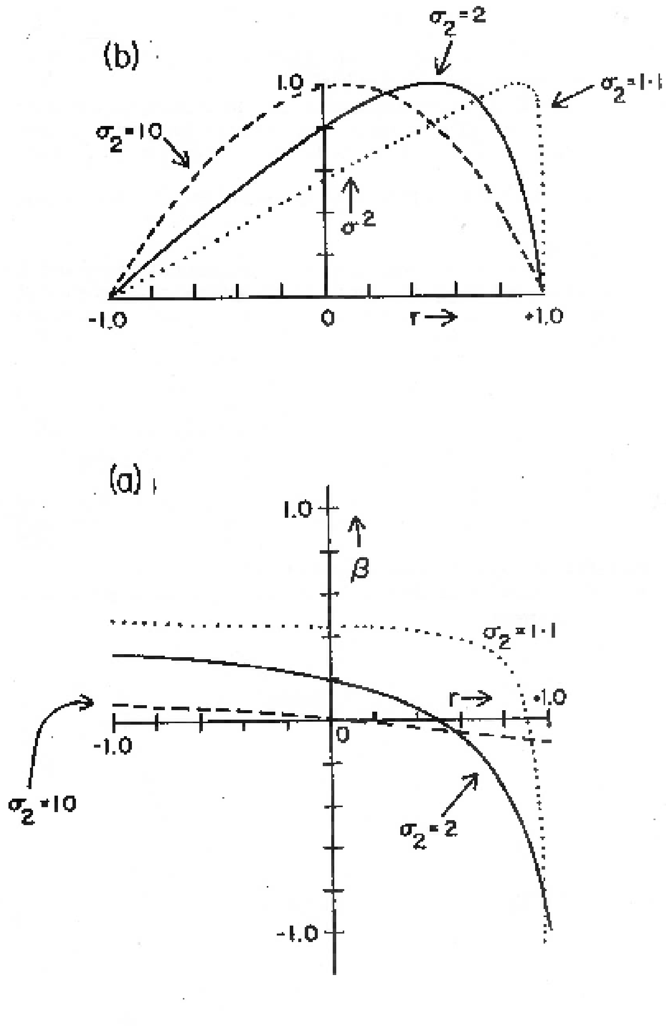


Fig. 1

Beware extrapolations because

[b] σ_{best} tends to zero, for $\rho = +1$ or -1

[a] v_{best} sensitive to ρ and σ_1/σ_2

N.B. For different analyses of ~ same data, $\rho \sim 1$, so choose 'better' analysis, rather than combining

N.B. σ_{best} depends on σ_1, σ_2 and ρ, but not on $v_1 - v_2$

e.g. Combining 0±3 and x±3 gives x/2 ± 2

BLUE = χ^2

$S(v_{best}) = \Sigma\, (v_i - v_{best})\, E^{-1}_{ij}\, (v_j - v_{best})$, and minimise S wrt v_{best}

S_{min} distributed like χ^2, so measures Goodness of Fit

But BLUE gives weights for each v_i

Can be used to see contributions to σ_{best} from each source of uncertainties e.g. statistical and systematics

different systematics

Extended by Valassi to combining more than one measured quantity e.g. intercepts and gradients of a straight line

MORE COMBINING:
Several pairs of correlated params

(x_i, y_i) with $\underline{\underline{E}}_i = \begin{pmatrix} \sigma_x^2 & cov \\ cov & \sigma_y^2 \end{pmatrix}_i$

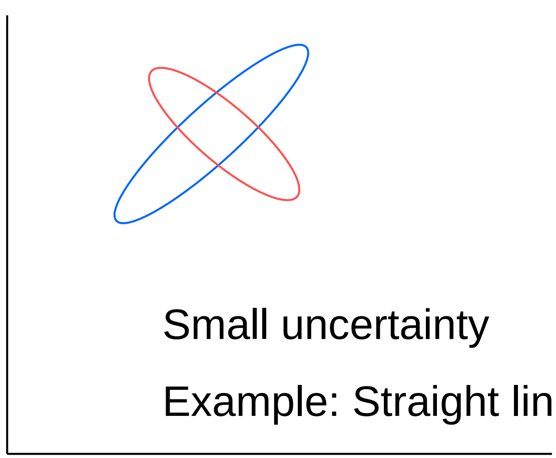
$$\chi^2 = \sum_i \left\{ (x_i - \hat{x})^2 E^{-1}_{11,i} + (y_i - \hat{y})^2 E^{-1}_{22,i} + 2(x_i - \hat{x})(y_i - \hat{y}) E^{-1}_{12,i} \right\}$$

i.e. result:—

Inverse error matrix on result $\hat{x}, \hat{y}$

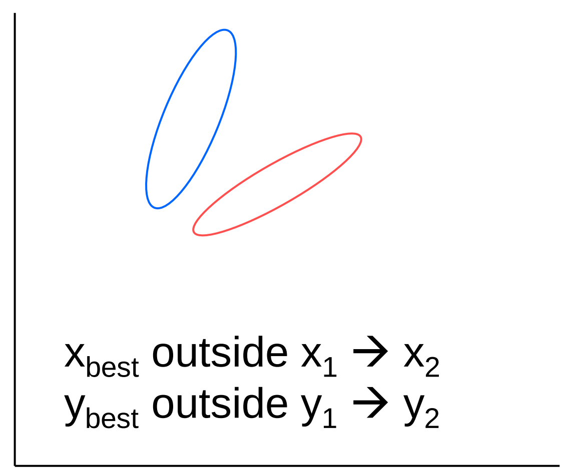
$$= \sum_i \underline{\underline{E}}_i^{-1}$$

cf $\frac{1}{\sigma^2} = \sum \frac{1}{\sigma_i^2}$ for single uncorrelated meas.



Small uncertainty

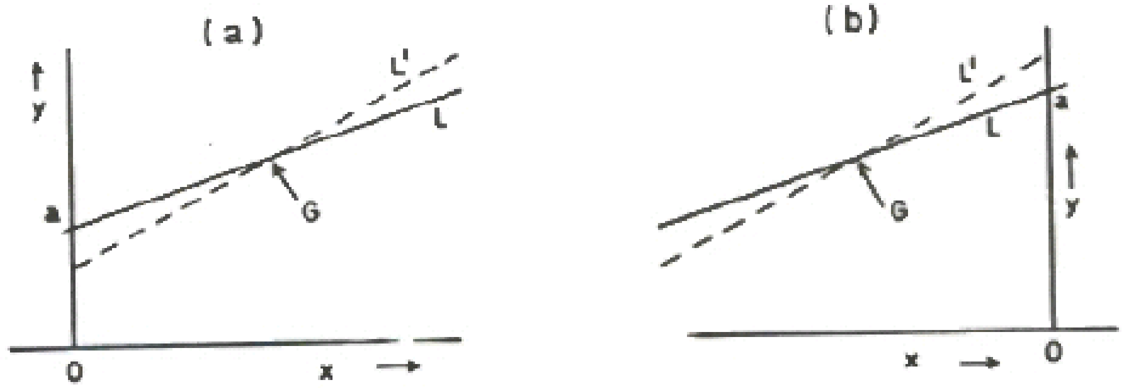
Example: Straight line fitting



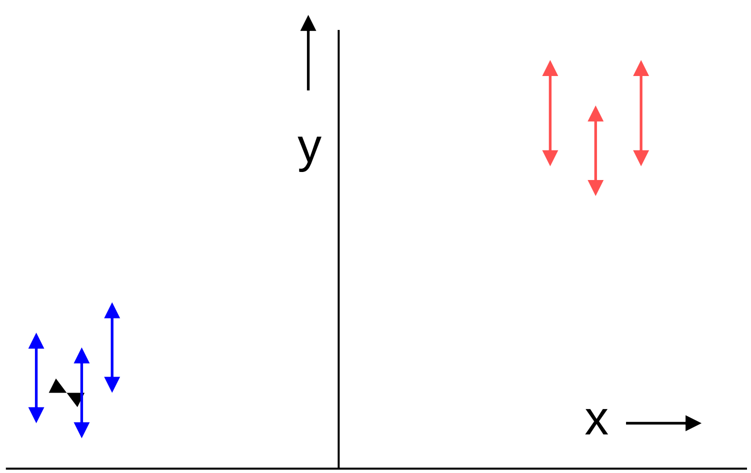
x_{best} outside $x_1 \rightarrow x_2$

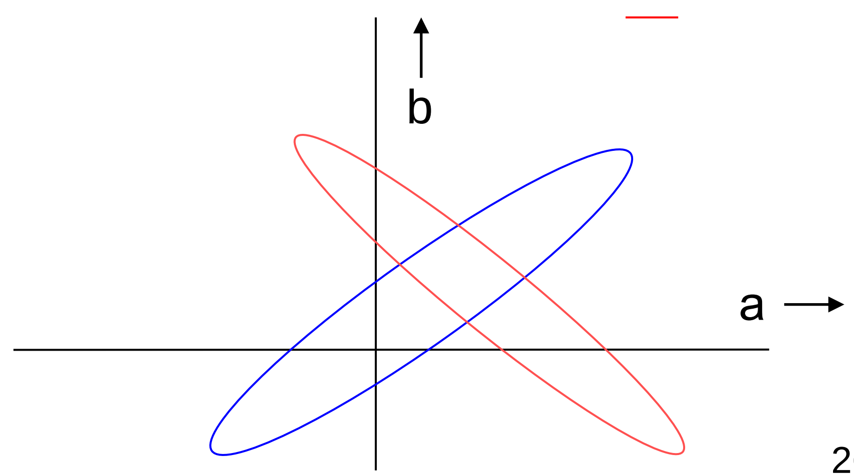
y_{best} outside $y_1 \rightarrow y_2$

COVARIANCE (a, b) $\propto -\langle x \rangle$



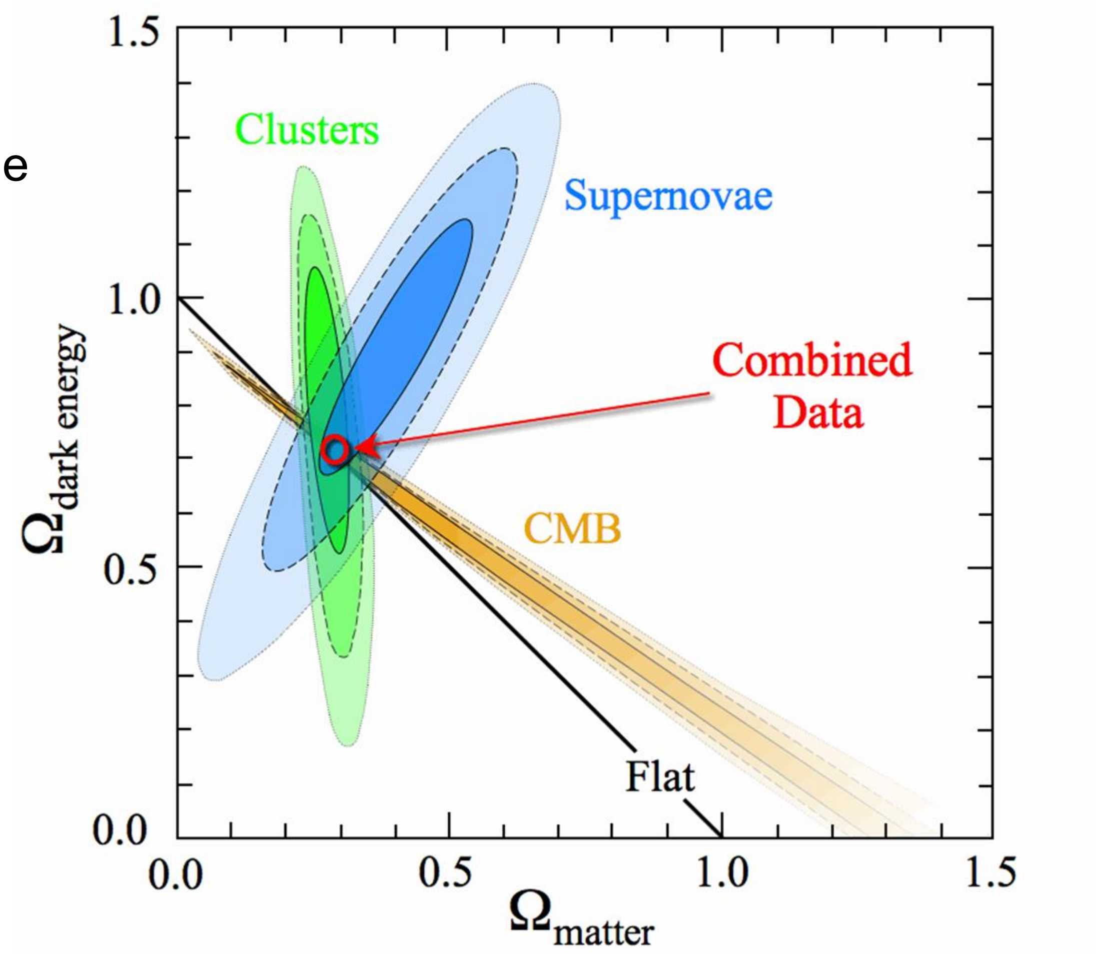
$\langle x \rangle$ pos   Fig. 2.4 $\langle x \rangle$ neg



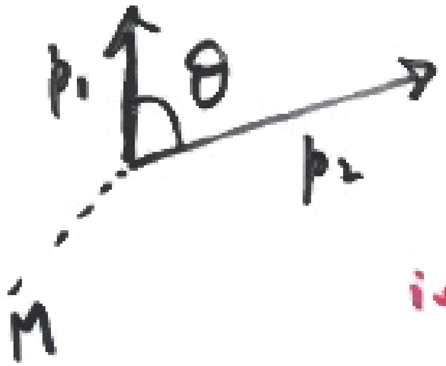


Uncertainty on $\Omega_{\text{dark energy}}$

When combining pairs of variables, the uncertainties on the combined parameters can be **much** smaller than any of the individual uncertainties
e.g. $\Omega_{\text{dark energy}}$



CORRELATIONS + MASS RESOLUTION



$$M^2 = (E_1 + E_2)^2 - (\underline{p}_1 + \underline{p}_2)^2$$

$$\sim p_1 p_2 \theta \qquad [\, p_i \gg m_i ; \ \theta \ll 1 \,]$$

i.e. $M \uparrow$ as $p_i \uparrow$ + $\theta_i \uparrow$

As $p_i \downarrow$, $\theta \uparrow$

$\therefore$ Smaller σ_M

As $p_i \downarrow$, $\theta \downarrow$

$\therefore$ Larger σ_M

Estimating the Covariance Matrix: $e^+ e^- \rightarrow W^+ W^-$

1) ESTIMATE ERRORS

ESTIMATE CORRELATIONS

(Usually easiest if $\rho = 0$ or ± 1)

2) FOR INDEP SOURCES OF ERRORS, ADD ERROR MATRICES

e.g. M_W FROM $WW \rightarrow$ 4 JETS

$WW \rightarrow JJ\ell\nu$

$\underline{\underline{E}}$ = $(M_W)_1, (M_W)_2$ ERROR MATRIX

$$\underline{\underline{E}} = \underline{\underline{E}}_{stat} + \underline{\underline{E}}_{B.E.} + \underline{\underline{E}}_{E\ scale} + \underline{\underline{E}}_{FSR} + \underline{\underline{E}}_{colour\ reconn}$$

Stat: $\begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$

B.E.: $\begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2 \\ \sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$

Colour reconn: $\begin{pmatrix} \sigma_1^2 & 0 \\ 0 & 0 \end{pmatrix}$

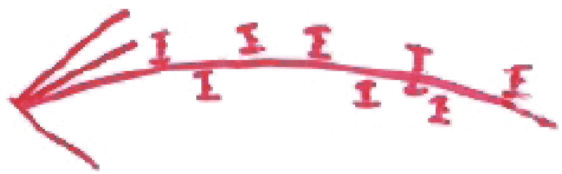
3) TRANSFORMATIONS

e.g. $(x \pm \sigma_x, y \pm \sigma_y)$ with <u>un</u>correl. errors

$\Rightarrow$ r, θ <u>with</u> correlations

Indep data points

$\Rightarrow$ correlated a and b



Track fit

4) REPEATED OBSERVATIONS

$(x_i, y_i) \Rightarrow \sigma_x^2 \quad \sigma_y^2$ and

$cov(x, y)$ from $\overline{(x-\bar{x})(y-\bar{y})}$

Conclusion

Covariance matrix formalism makes life easy when correlations are relevant