For general linear regression, the chi-square is explicitly quadratic in the parameters:

$$\chi^2 = (y - A\theta)^T\Sigma^{-1}(y - A\theta)$$

In particular, only one (global) minimum

For the general minimization problem, there can be local and global minima

But even then, the chi-square can be expanded around the minimum in a multi-dimensional second order polynomial → all that follows about parameter uncertainties (fluctuations around the minimum) is applicable to the general problem.
Parameter errors

- Consider fit hypothesis of last exercise of Exercise Sheet 2:
  \[ y = a + bx \quad (a^0 = 1, \ b^0 = 0.5) \]

- True parameter values are known in this case. One of the tasks was to save a list with the parameter estimates \( \hat{a}, \hat{b} \) for every fit.

- These estimates represent the minimal chi-square for a given data set. Repeating these fits for many data sets (all randomly distributed around \( y = a^0 + b^0 x \) ) resulted in this list. This is a scatter plot of your results:

Observations:
- Parameter estimates are indeed distributed around (1.0,0.5)
- There is a spread as expected from the statistical nature of the ensemble data
- There is, moreover, an additional structure: Combinations of large \( a \) with small \( b \) are favored → Parameters are anti-correlated.
Uncertainty for a one-parameter fit

- Two challenges:
  - What is the spread in a parameter?
  - That spread depends on the consider parameter combination itself → “Correlations”

- Question one: for one parameter, what is its spread?

Simple example: consider the fit function $y=a$. We know already: $\hat{a}$ will be the estimate of the mean and its uncertainty is

$$(\Delta a)^2 = E[(\hat{a} - a^0)^2] = \frac{\sigma^2}{n}$$

where $\sigma$ is the standard deviation of the data itself. This answers our question regarding the parameter uncertainty, but we would like to derive the parameter uncertainty from the behavior of the chi-square function as a function of $a$. 
Show (blackboard): \[ E[\chi^2(a)] = n + (a - a_0) \frac{1}{(\Delta a)^2} (a - a_0) \]

Means that \( \Delta a \) can be determined by determining the \( a \) for which chi-square rises by +1, compared to the best chi-square:

\[
E[\chi^2(a_0 + \Delta a)] = n + (\Delta a) \frac{1}{(\Delta a)^2} (\Delta a) = n + 1
\]

\( \Delta a \) is normal distributed because the \( y_i \) are: \( \hat{a} = \frac{1}{n} \sum_{i=1}^{n} y_i \)

In fact, as we know, \( \frac{\hat{a} - a^0}{\sigma/\sqrt{n}} \) is standard Normal distributed, \( \mathcal{N}[0,1] \).

Suggested generalization to many parameters:

\[
E[\chi^2(a)] = \chi^2(a_0) + (a - a_0)^T C^{-1} (a - a_0), \quad C = \begin{pmatrix}
(\Delta a)^2 & ? & ? \\
? & (\Delta b)^2 & ? \\
\vdots & \vdots & \ddots
\end{pmatrix}
\]

C: “Correlation matrix”
One step in previous derivation was that parameter uncertainties can be derived from the second derivative of the chi-square function, as a function of the parameters. Careful: linear error ≠ non-linear error. Consider extreme case:

- In general: Asymmetric error bars
- Errors cannot be determined from 2nd derivatives, but only with numerical simulations ("Monte-Carlo", "Bootstrap")
- → See exercises (you are experts now)!

Note: all quantities carry “^” because in practical problems that is all there is (a₀, x₀^2 unknown). Uncertainties themselves are estimates.
Bottom line

- For many practical purposes, we can think of parameter uncertainties as if they were from a general linear regression, that has a chi-square quadratic in the parameters.

- This is because locally, the chi-square can be expanded around the minimum, even in the general chi-square problem.

- For the general chi-square problem, the situation will be similar, except for pathological situations (that may not be so pathological after all). Keep this caveat in mind.
Uncertainties for several parameters

- We guessed a suitable form
  \[ E[\chi^2(\theta)] = \chi^2(\theta_0) + (\theta - \theta_0)^T C^{-1} (\theta - \theta_0) \]

- Show (for general linear regression):
  \[ \hat{\chi}^2(\theta) = \chi^2(\hat{\theta}) + (\theta - \hat{\theta})^T C^{-1} (\theta - \hat{\theta}), \quad C^{-1} = A^T \Sigma^{-1} A \]

  Here, \( \Sigma \) is the diagonal matrix of data uncertainties (blackboard). (Note: notation updated to exclude confusion; please update your handwritten notes)

- Although we can calculate C in general linear regression, we cannot yet connect it to parameter errors. Instead, we have to look first what distribution the parameters follow. That distribution will have (co)variances, from which we can read off the parameter uncertainties. And it will contain C such that we can finally relate C to uncertainties.