General Least Squares with General Constraints:

Suppose we have P ***measurements*** $\underline{y}(\underline{x}) \equiv (y_1(x_1), y_2(x_2), \ldots, y_P(x_P))$ with a ***symmetric*** $P \times P$ ***covariance*** matrix of the $\underline{y}(\underline{x})$ ***measurements*** $\underline{V}_{\underline{y}(\underline{x})}$. Suppose the ***theory prediction*** $\underline{\bar{y}}(\underline{x};\underline{\lambda}) \equiv (\bar{y}_1(x_1;\underline{\lambda}), \bar{y}_2(x_2;\underline{\lambda}), \ldots, \bar{y}_P(x_P;\underline{\lambda}))$ involves M $(< P)$ ***parameters*** $\underline{\lambda} \equiv (\lambda_1, \lambda_2, \ldots, \lambda_M)$ in some ***general*** (*i.e.* not necessarily ***linear***) manner. Additionally, suppose there are K ***functions*** $\underline{f}(\underline{\lambda}) \equiv (f_1(\underline{\lambda}), f_2(\underline{\lambda}), \ldots, f_K(\underline{\lambda}))$ that relate (*i.e.* ***constrain***) the M $\underline{\lambda}$***-parameters*** in some ***general*** (but not necessarily ***linear*** manner) via use of ***Lagrange Multipliers*** $\underline{\alpha} \equiv (\alpha_1, \alpha_2, \ldots, \alpha_K)$.

The $\chi^2(\underline{\lambda};\underline{\alpha})$ is defined as:

$$\underbrace{\chi^2(\underline{\lambda};\underline{\alpha})}_{1\times1} \equiv \underbrace{\chi^2(\underline{\lambda})}_{1\times1} + 2\underbrace{\underline{\alpha}^T}_{1\times K}\underbrace{\underline{f}(\underline{\lambda})}_{K\times1} = \underbrace{\left(\underline{y}(\underline{x}) - \underline{\bar{y}}(\underline{x};\underline{\lambda})\right)^T}_{1\times P}\underbrace{\underline{V}^{-1}_{\underline{y}(\underline{x})}}_{P\times P}\underbrace{\left(\underline{y}(\underline{x}) - \underline{\bar{y}}(\underline{x};\underline{\lambda})\right)}_{1\times P} + 2\underbrace{\underline{\alpha}^T}_{1\times K}\underbrace{\underline{f}(\underline{\lambda})}_{K\times1}$$

where $\underline{V}^{-1}_{\underline{y}(\underline{x})}$ is the $P \times P$ ***symmetric inverse*** of the ***covariance*** matrix of the $\underline{y}(\underline{x})$ ***measurements***, and the $K \times 1$ column vector $\underline{f}(\underline{\lambda})$ contains the K ***constraint*** equations. {*n.b.* In the ***linear constraint*** case $\underline{f}(\underline{\lambda}) = \underline{B}\underline{\lambda} - \underline{b}$. However, in ***general*** the ***constraint*** equations $\underline{f}(\underline{\lambda})$ ***may*** be ***non-linear*** functions of the M $\underline{\lambda}$***-parameters***.}

We minimize the $\chi^2(\underline{\lambda};\underline{\alpha})$ by taking derivatives *w.r.t.* $(\underline{\lambda};\underline{\alpha})$. We (again) use the iteration technique here too. Suppose that after ν iterations, we have obtained a set of ***approximate*** values of the M $\underline{\lambda}$***-parameters*** and K ***Lagrange Multipliers*** $\underline{\alpha}$:

$$\underline{\lambda}^\nu = \begin{pmatrix} \lambda_1^\nu \\ \lambda_2^\nu \\ \vdots \\ \lambda_M^\nu \end{pmatrix} \quad \text{and:} \quad \underline{\alpha}^\nu = \begin{pmatrix} \alpha_1^\nu \\ \alpha_2^\nu \\ \vdots \\ \alpha_K^\nu \end{pmatrix}$$

We then expand (*i.e.* ***linearize***) $\chi^2(\underline{\lambda};\underline{\alpha})$ in a Taylor series around these points $(\underline{\lambda}^\nu;\underline{\alpha}^\nu)$, then solve for $\Delta\underline{\lambda}^\nu \equiv (\Delta\lambda_1^\nu, \Delta\lambda_2^\nu, \ldots, \Delta\lambda_M^\nu)$, $\Delta\underline{\alpha}^\nu \equiv (\Delta\alpha_1^\nu, \Delta\alpha_2^\nu, \ldots, \Delta\alpha_K^\nu)$ and iterate further – similar to the discussion in P598AEM Lect. Notes 20 (*p.* 5-9). For additional details, see *e.g.* individual program write-ups or *e.g.* advanced texts on this subject...

Let us assume that we have determined the "***best***" values $(\underline{\lambda}^*;\underline{\alpha}^*)$ of these parameters using the Lagrange Multiplier constrained LSQ fit method.

We can obtain a ***better*** estimate, if we wish, of the ***measured*** ***random variables*** $\underline{y}(\underline{x})$. This procedure goes by the name "***Adjustment of Observations***":

We define a $P\times 1$ column vector $\underline{m} \equiv \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_P \end{pmatrix}$ of ***measured*** values of the P ***random variables*** (*n.b.* these may not necessarily be ***independent***), with corresponding $P\times P$ ***symmetric covariance*** matrix $\underline{V}_{\underline{m}}$ of the ***measurements*** $\underline{m}$.

We want to know the "***true***" values (*i.e.* ***expectation values***) of the ***measurements***:

$$E[\underline{m}] = \underline{\hat{m}} \equiv (\hat{m}_1, \hat{m}_2, \ldots, \hat{m}_P).$$

We will ***estimate*** them using a LSQ fitting method, and call the ***estimates*** the "***fitted values of the measurements***". We obtain the "***fitted values of the measurements***" by ***adjusting*** the ***measurements*** so that:

- Each measurement m_i is allowed to move by an amount determined from the size of the ***uncertainty*** on the measurement, σ_{m_i}.
- The resulting ***fitted values of the measurements*** satisfy one or more ***constraints***.

We define a $P\times 1$ column vector: $\underline{f} \equiv \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_P \end{pmatrix}$ of ***fitted values*** of $\underline{m}$, *i.e.* the ***estimates*** of $\underline{\hat{m}}$.

Let there be K ***constraints*** which can be expressed in the form:

$$\left\{\begin{matrix} \mathbb{C}_1(f_1, f_2, \ldots, f_P) = 0 \\ \mathbb{C}_2(f_1, f_2, \ldots, f_P) = 0 \\ \vdots \\ \mathbb{C}_K(f_1, f_2, \ldots, f_P) = 0 \end{matrix}\right\} \text{ or, defining a } K\times 1 \text{ column vector: } \underline{\mathbb{C}}(\underline{f}) \equiv \begin{pmatrix} \mathbb{C}_1(f_1, f_2, \ldots, f_P) \\ \mathbb{C}_2(f_1, f_2, \ldots, f_P) \\ \vdots \\ \mathbb{C}_K(f_1, f_2, \ldots, f_P) \end{pmatrix} = 0$$

n.b. In ***general***, these will be ***non-linear*** equations.

Remembering the ***iterative*** χ^2 ***minimization*** method(s), we choose to work with ***linearized*** "***corrections***":

$$\left\{\begin{matrix} c_1 = f_1 - m_1 \\ c_2 = f_2 - m_2 \\ \vdots \\ c_P = f_P - m_P \end{matrix}\right\} \text{ or, defining a } P\times 1 \text{ column vector: } \underline{c} \equiv \underline{f} - \underline{m}$$

In terms of χ^2 ***minimization***, since the $\underline{m}$'s are just ***constants***, ***minimizing*** χ^2 with respect to $\underline{c} \equiv \underline{f} - \underline{m}$ is equivalent to ***minimizing*** χ^2 with respect to $\underline{f}$.

What should we ***actually minimize***? If we use $\chi^2 \equiv \left(\underline{f} - \underline{m}\right)^T \underline{V}_{\underline{m}}^{-1} \left(\underline{f} - \underline{m}\right) = \underline{c}^T \underline{V}_{\underline{m}}^{-1} \underline{c} = \chi^2\left(\underline{c}\right)$, the solution is (obviously) $\underline{f} = \underline{m}$, *i.e.* the "***best***" ***estimate*** of $\hat{m}_i$ is m_i itself. In order to do ***better***, we must add in some ***new*** information – in this case, the requirement that the ***constraints*** be satisfied by the $\underline{f}$'s. Thus, we instead minimize:

$$\begin{aligned}\chi^2\left(\underline{c};\underline{\alpha}\right) &\equiv \chi^2 + 2\underline{\mathbb{C}}^T\underline{\alpha} \\ &= \underline{c}^T \underline{V}_{\underline{m}}^{-1} \underline{c} + 2\underline{\mathbb{C}}^T\underline{\alpha}\end{aligned} \quad \text{where: } \underline{\alpha} \equiv \begin{pmatrix}\alpha_1\\ \alpha_2\\ \vdots \\ \alpha_K\end{pmatrix}$$

is a $K \times 1$ column vector of ***Lagrange Multipliers***.

Taking derivatives of $\chi^2\left(\underline{c};\underline{\alpha}\right)$, we obtain:

$$0 = \frac{\partial \chi^2\left(\underline{c};\underline{\alpha}\right)}{\partial \underline{\alpha}} \Rightarrow \underline{\mathbb{C}}^T\left(\underline{f}\right) = 0 \quad (\textit{i.e.} \text{ the constraints } \underline{\textbf{\textit{will}}} \text{ be satisfied})$$

$$0 = \frac{\partial \chi^2\left(\underline{c};\underline{\alpha}\right)}{\partial \underline{c}} = \frac{\partial \chi^2\left(\underline{f} - \underline{m};\underline{\alpha}\right)}{\partial \underline{f}} \quad \Rightarrow \quad \underline{V}_{\underline{m}}^{-1}\underline{c} + \frac{\partial \underline{\mathbb{C}}^T\left(\underline{f}\right)}{\partial \underline{f}}\underline{\alpha} = 0$$

Note that $\dfrac{\partial \underline{\mathbb{C}}^T\left(\underline{f}\right)}{\partial \underline{f}}$ is a $P \times K$ matrix $\left(\equiv \underline{B}\left(\underline{f}\right)\right)$ with jk^{th} element: $B_{jk} = \dfrac{\partial \mathbb{C}_k\left(\underline{f}\right)}{\partial f_j}$, where $j = 1, 2, \ldots, P$ ranges over the ***fitted*** variables and $k = 1, 2, \ldots, K$ ranges over the ***constraints***.

Thus, the equations that we need to solve in order to accomplish this $\chi^2\left(\underline{c};\underline{\alpha}\right)$ minimization are:

$$\underbrace{\underline{\mathbb{C}}^T\left(\underline{f}\right)}_{1\times K} = 0 \quad \text{and:} \quad \underbrace{\underline{V}_{\underline{m}}^{-1}}_{P\times P}\underbrace{\underline{c}}_{P\times 1} + \underbrace{\underline{B}\left(\underline{f}\right)}_{P\times K}\underbrace{\underline{\alpha}}_{K\times 1} = 0$$

For the ***general*** ***non-linear*** case, we must resort to approximation methods. We Taylor series expand (*i.e.* ***linearize***) the ***constraint*** equations around $\underline{f}^0$, an ***initial*** ***estimate*** of the ***fitted values of the measurements*** $\underline{f}$. Then we require that:

$$\underbrace{\underline{\mathbb{C}}\left(\underline{f}\right)}_{K\times 1} = 0 = \underbrace{\underline{\mathbb{C}}\left(\underline{f}^0\right)}_{K\times 1} + \underbrace{\left[\frac{\partial \underline{\mathbb{C}}\left(\underline{f}\right)}{\partial \underline{f}}\right]_{\underline{f}=\underline{f}^0}}_{K\times P} \underbrace{\left(\underline{f} - \underline{f}^0\right)}_{P\times 1} + \cdots$$

As usual, we ***assume*** that $\underline{f} - \underline{f}^0$ is ***small*** ***enough*** so that we can ***safely*** neglect/ignore the terms in the Taylor series expansion involving higher powers of $\left(\underline{f} - \underline{f}^0\right)$ ***.and.*** the higher-order derivatives of $\underline{\mathbb{C}}\left(\underline{f}\right)$. (This step is known as "***linearizing*** the ***constraints***".)

Then:

$$0 = \underline{\mathbb{C}}\left(\underline{f}^0\right) + \left[\frac{\partial\underline{\mathbb{C}}\left(\underline{f}\right)}{\partial\underline{f}}\right]_{\underline{f}=\underline{f}^0}\left(\underline{f} - \underline{f}^0\right) = \underline{\mathbb{C}}\left(\underline{f}^0\right) + \underline{B}^T_{\underline{f}=\underline{f}^0}\left(\underline{f} - \underline{f}^0\right)$$

A neat trick exists for solving this conveniently. We write:

$$\left(\underline{f} - \underline{f}^0\right) \equiv \left(\underline{f} - \underline{m}\right) - \left(\underline{f}^0 - \underline{m}\right) \equiv \left(\underline{c} - \underline{c}^0\right)$$

where: $\underline{c} \equiv \underline{f} - \underline{m}$ and: $\underline{c}^0 \equiv \left(\underline{f}^0 - \underline{m}\right)$

Then: $$0 = \underline{\mathbb{C}}\left(\underline{f}^0\right) + \left[\frac{\partial\underline{\mathbb{C}}\left(\underline{f}\right)}{\partial\underline{f}}\right]_{\underline{f}=\underline{f}^0}\left(\underline{c} - \underline{c}^0\right) = \underline{\mathbb{C}}\left(\underline{f}^0\right) + \underline{B}^T_{\underline{f}=\underline{f}^0}\left(\underline{c} - \underline{c}^0\right)$$

We rewrite this as: $0 = \underline{\mathbb{C}} + \underline{B}^T\left(\underline{c} - \underline{c}^0\right)$ where it is implicitly understood that the derivatives and the constraints are evaluated at the an ***initial*** ***<u>estimate</u>*** $\underline{f} = \underline{f}^0$. Then: $\boxed{\underline{B}^T\underline{c} = \underline{B}^T\underline{c}^0 - \underline{\mathbb{C}} \equiv \underline{r}}$.

The other equation we must solve is: $\underline{V}_{\underline{m}}^{-1}\underline{c} + \underline{B}\underline{\alpha} = 0$, which, multiplying on the LHS by $\underline{V}_{\underline{m}}$ yields: $\boxed{\underline{c} = -\underline{V}_{\underline{m}}\underline{B}\underline{\alpha}}$.

Thus: $\underline{r} = \underline{B}^T\underline{c} = -\left(\underline{B}^T\underline{V}_{\underline{m}}\underline{B}\right)\underline{\alpha} \equiv -\underline{H}\,\underline{\alpha}$ where: $\underbrace{\underline{H}}_{K\times K} \equiv \underbrace{\underline{B}^T}_{K\times P}\underbrace{\underline{V}_{\underline{m}}}_{P\times P}\underbrace{\underline{B}}_{P\times K}$

By construction, $\underline{H} \equiv \underline{B}^T\underline{V}_{\underline{m}}\underline{B}$ is a $K \times K$ square, ***symmetric*** (and ***real***) matrix, and therefore, it has a $K \times K$ square, ***symmetric*** (and ***real***) ***<u>inverse</u>*** $\underline{H}^{-1} \equiv \left(\underline{B}^T\underline{V}_{\underline{m}}\underline{B}\right)^{-1}$.

Thus, multiplying $\underline{r} = \underline{B}^T\underline{c} = -\left(\underline{B}^T\underline{V}_{\underline{m}}\underline{B}\right)\underline{\alpha} \equiv -\underline{H}\,\underline{\alpha}$ on the LHS by $\underline{H}^{-1} \equiv \left(\underline{B}^T\underline{V}_{\underline{m}}\underline{B}\right)^{-1}$ gives the ***Lagrange Multipliers***: $\boxed{\underline{\alpha} = -\underline{H}^{-1}\underline{r}}$ and $\boxed{\underline{c} = -\underline{V}_{\underline{m}}\underline{B}\underline{\alpha} = +\underline{V}_{\underline{m}}\underline{B}\underline{H}^{-1}\underline{r}}$ gives the "***correction***".

Finally, the ***<u>result</u>*** of this step is: $\boxed{\underline{f} = \underline{m} + \underline{c}}$.

We ***explicitly*** need to check/verify whether or not this new $\underline{f}$ satisfies the ***constraints***: $\underline{\mathbb{C}}\left(\underline{f}\right) = 0$.

If it does, then we're done. If not, then we use ***<u>this</u>*** $\underline{f}$ as a ***<u>new</u>*** $\underline{f}^0$ and repeat (*i.e.* ***<u>iterate</u>***) the above procedure until $\underline{\mathbb{C}}\left(\underline{f}\right) = 0$ ***<u>is</u>*** satisfied. If $\underline{\mathbb{C}}\left(\underline{f}\right) = 0$ ***<u>is</u>*** satisfied, then $\underline{\mathbb{C}}^T\left(\underline{f}\right) = 0$ also.

Now let us calculate $\chi^2(\underline{c};\underline{\alpha})$ from the quantities that we have obtained.

If $\underline{\mathbb{C}}^T(\underline{f})=0$ ***is*** satisfied, recalling that $\underline{V}_{\underline{m}}$ and $\underline{H}\equiv\underline{B}^T\underline{V}_{\underline{m}}\underline{B}$ are ***symmetric*** matrices, then:

$$
\begin{aligned}
\chi^2(\underline{c};\underline{\alpha}) &= \underline{c}^T\underline{V}_{\underline{m}}^{-1}\underline{c}+2\underline{\mathbb{C}}^T\underline{\alpha}=\underline{c}^T\underline{V}_{\underline{m}}^{-1}\underline{c}\\
&=\left(\underline{V}_{\underline{m}}\underline{B}\underline{H}^{-1}\underline{r}\right)^T\underline{V}_{\underline{m}}^{-1}\left(\underline{V}_{\underline{m}}\underline{B}\underline{H}^{-1}\underline{r}\right)\\
&=\underline{r}^T\underline{H}^{-1}\underline{B}^T\,\cancel{\underline{V}_{\underline{m}}\underline{V}_{\underline{m}}^{-1}}\left(\underline{V}_{\underline{m}}\underline{B}\underline{H}^{-1}\underline{r}\right)\\
&=\underline{r}^T\underline{H}^{-1}\underbrace{\left(\underline{B}^T\underline{V}_{\underline{m}}\underline{B}\right)}_{\equiv\underline{H}}\underline{H}^{-1}\underline{r}\\
&=\underline{r}^T\,\cancel{\underline{H}^{-1}\underline{H}}\,\underline{H}^{-1}\underline{r}\\
&=\underline{r}^T\underbrace{\underline{H}^{-1}\underline{r}}_{=-\underline{\alpha}}\\
&=-\underline{r}^T\underline{\alpha}
\end{aligned}
$$

Thus, $\chi^2(\underline{c};\underline{\alpha})=-\underline{r}^T\underline{\alpha}$. This is the value of χ^2 ***after*** the step to $\underline{f}=\underline{m}+\underline{c}$.

Next, we determine the $P\times P$ ***covariance*** matrix of the "***fitted values***" using ***error propagation***:

$$
\underbrace{\underline{V}_{\underline{f}}}_{P\times P}\equiv\underbrace{\left(\frac{\partial\underline{f}}{\partial\underline{m}}\right)}_{P\times P}\underbrace{\underline{V}_{\underline{m}}}_{P\times P}\underbrace{\left(\frac{\partial\underline{f}}{\partial\underline{m}}\right)^T}_{P\times P}
$$

Now it is just algebra...

$$
\frac{\partial\underline{f}}{\partial\underline{m}}=\frac{\partial(\underline{m}+\underline{c})}{\partial\underline{m}}=\underline{\mathbb{1}}+\frac{\partial\underline{c}}{\partial\underline{m}}=\underline{\mathbb{1}}+\frac{\partial\left(\underline{V}_{\underline{m}}\underline{B}\underline{H}^{-1}\underline{r}\right)}{\partial\underline{m}}=\underline{\mathbb{1}}+\underline{V}_{\underline{m}}\underline{B}\underline{H}^{-1}\frac{\partial\underline{r}}{\partial\underline{m}}
$$

But: $\underline{r}=\underline{B}^T\underline{c}=\underline{B}^T(\underline{f}-\underline{m})$, thus: $\dfrac{\partial\underline{r}}{\partial\underline{m}}=\dfrac{\partial\underline{B}^T(\underline{f}-\underline{m})}{\partial\underline{m}}=-\underline{B}^T\underline{\mathbb{1}}=-\underline{B}^T$ and thus: $\boxed{\dfrac{\partial\underline{f}}{\partial\underline{m}}=\underline{\mathbb{1}}-\underline{V}_{\underline{m}}\left(\underline{B}\underline{H}^{-1}\underline{B}^T\right)}$

Then:

$$
\begin{aligned}
\underline{V}_{\underline{f}}\equiv\left(\frac{\partial\underline{f}}{\partial\underline{m}}\right)\underline{V}_{\underline{m}}\left(\frac{\partial\underline{f}}{\partial\underline{m}}\right)^T &=\left(\underline{\mathbb{1}}-\underline{V}_{\underline{m}}\left(\underline{B}\underline{H}^{-1}\underline{B}^T\right)\right)\underline{V}_{\underline{m}}\left(\underline{\mathbb{1}}-\underline{V}_{\underline{m}}\left(\underline{B}\underline{H}^{-1}\underline{B}^T\right)\right)^T\\
&=\left(\underline{\mathbb{1}}-\underline{V}_{\underline{m}}\left(\underline{B}\underline{H}^{-1}\underline{B}^T\right)\right)\underline{V}_{\underline{m}}\left(\underline{\mathbb{1}}-\left(\underline{B}\underline{H}^{-1}\underline{B}^T\right)\underline{V}_{\underline{m}}\right)^T
\end{aligned}
$$

Multiplying this out on the RHS and again using $\underline{H}\equiv\underline{B}^T\underline{V}_{\underline{m}}\underline{B}$ this simplifies to:

$$
\boxed{\underline{V}_{\underline{f}}=\underline{V}_{\underline{m}}-\left(\underline{V}_{\underline{m}}\underline{B}\right)\underline{H}^{-1}\left(\underline{V}_{\underline{m}}\underline{B}\right)^T}
$$

As before, since $\underline{\underline{V}}_m$ is ***symmetric***, it has ***positive diagonal*** elements. Likewise, the ***symmetric*** matrix $\underline{\underline{H}} \equiv \underline{\underline{B}}^T \underline{\underline{V}}_m \underline{\underline{B}}$ ***also*** has ***positive diagonal*** elements, and so does $\underline{\underline{H}}^{-1} \equiv \left(\underline{\underline{B}}^T \underline{\underline{V}}_m \underline{\underline{B}}\right)^{-1}$.

Therefore, from $\underline{\underline{V}}_f = \underline{\underline{V}}_m - \left(\underline{\underline{V}}_m \underline{\underline{B}}\right) \underline{\underline{H}}^{-1} \left(\underline{\underline{V}}_m \underline{\underline{B}}\right)^T = \underline{\underline{V}}_m - \left(\underline{\underline{V}}_m \underline{\underline{B}}\right)\left(\underline{\underline{B}}^T \underline{\underline{V}}_m \underline{\underline{B}}\right)^{-1}\left(\underline{\underline{V}}_m \underline{\underline{B}}\right)^T$, we see that the ***diagonal*** elements of $\underline{\underline{V}}_f$ are ***smaller*** than the diagonal elements of $\underline{\underline{V}}_m$.

Thus, the 1-standard deviation uncertainties associated the ***adjusted*** (*i.e.* ***fitted***) ***measurements*** are less than the 1-standard deviation uncertainties on the ***original measurements***.

"Pull" Quantities:

"Pull" quantities are distributions of ***normalized/fractional differences*** between the ***fitted – measured*** quantities. which can be very helpful in verifying the validity of the LSQ fitting procedure.

We define the i^{th} "pull" quantity as the ***normalized correction***:

$$p_i \equiv \frac{c_i}{\sqrt{\left\langle c_i^2 \right\rangle}} = \frac{f_i - m_i}{\sqrt{\left\langle \left(f_i - m_i\right)^2 \right\rangle}}$$

where the brackets $\langle\ \rangle$ are ***synonymous*** with the ***expectation value***, *i.e.*:

$$\left\langle c_i^2 \right\rangle = E[c_i^2] \;\Rightarrow\; \left\langle \left(f_i - m_i\right)^2 \right\rangle = E[\left(f_i - m_i\right)^2]$$

Note that if there is ***no bias***, then: $\langle c_i \rangle = \langle f_i - m_i \rangle = 0$. If everything is "nice" – *i.e.* the P input ***measurements*** $\underline{m}$ are Gaussian/normally-distributed ***.and.*** their ***uncertainties***, as contained in the individual elements of the $P \times P$ ***covariance*** matrix of the ***measurements*** $\underline{\underline{V}}_m$ have ***all*** been correctly / properly assigned ***.and.*** if the various approximations and assumptions are ***all*** valid, then the p_i ***should*** be distributed as $N(0,1)$.

By explicitly looking at the distributions (*e.g.* histograms) of the p_i for many ***independent measurements*** of each of the m_i, we can turn this around and check the "ingredients" listed above, especially whether the ***uncertainties*** on the individual have indeed been correctly assigned or not, by seeing whether the "pull" distribution p_i for each m_i ***is*** indeed distributed as $N(0,1)$ or not.

Let us suppose that we have performed the ***Adjustment of Observations***, starting with our ***initial measurements*** $\underline{m}$ and arriving at ***final adjusted/fitted values*** $\underline{f}$.

It is not trivial to evaluate the $\left\langle c_i^2 \right\rangle$. The $P \times 1$ column vector ***correction*** $\underline{c} \equiv \underline{f} - \underline{m}$. We also have the $P \times P$ ***covariance*** matrix of the ***measurements*** $\underline{\underline{V}}_m$ and that of the ***adjusted/fitted measurements*** $\underline{\underline{V}}_f$.

Formally: $$\left(\underline{\underline{V}}_m\right)_{ij} = E[(m_i - \hat{m}_i)(m_j - \hat{m}_j)] = \left\langle (m_i - \hat{m}_i)(m_j - \hat{m}_j) \right\rangle$$

and: $$\left(\underline{\underline{V}}_f\right)_{ij} = E[(f_i - \hat{f}_i)(f_j - \hat{f}_j)] = \left\langle (f_i - \hat{f}_i)(f_j - \hat{f}_j) \right\rangle$$

If the P ***measurements*** $\underline{m}$ are truly ***unbiased***, then: $\hat{m}_i = \hat{f}_i$, *i.e.* $E[m_i] = \langle m_i \rangle = \langle f_i \rangle = E[f_i]$.

Thus: $$c_i \equiv f_i - m_i = \left(f_i - \hat{f}_i\right) - \left(m_i - \hat{m}_i\right)$$

For convenience, we define the $P \times 1$ column vectors: $\underline{\delta f} \equiv \left(\underline{f} - \underline{\hat{f}}\right)$ and: $\underline{\delta m} \equiv \left(\underline{m} - \underline{\hat{m}}\right)$.

Then: $$\underline{c} \equiv \underline{f} - \underline{m} = \left(\underline{f} - \underline{\hat{f}}\right) - \left(\underline{m} - \underline{\hat{m}}\right) = \underline{\delta f} - \underline{\delta m} \quad \text{or:} \quad \underline{\delta m} = \underline{\delta f} - \underline{c}$$

and:

$$\underline{V}_{\underline{m}} = E[\underline{\delta m}\,\underline{\delta m}^T] = \left\langle \underline{\delta m}\,\underline{\delta m}^T \right\rangle = \left\langle \left(\underline{\delta f} - \underline{c}\right)\left(\underline{\delta f} - \underline{c}\right)^T \right\rangle$$
$$= \left\langle \underline{\delta f}\,\underline{\delta f}^T \right\rangle + \left\langle \underline{c}\,\underline{c}^T \right\rangle - 2\left\langle \underline{c}\,\underline{\delta f}^T \right\rangle$$

or: $$\underline{V}_{\underline{m}} = \underline{V}_{\underline{f}} + \left\langle \underline{c}\,\underline{c}^T \right\rangle - 2\left\langle \underline{c}\,\underline{\delta f}^T \right\rangle$$

or: $$\underline{V}_{\underline{c}} \equiv \left\langle \underline{c}\,\underline{c}^T \right\rangle = \underline{V}_{\underline{m}} - \underline{V}_{\underline{f}} + 2\left\langle \underline{c}\,\underline{\delta f}^T \right\rangle$$

This ***is*** what we need, since the ***diagonal*** elements of the $P \times P$ ***covariance*** matrix $\underline{V}_{\underline{c}} \equiv \left\langle \underline{c}\,\underline{c}^T \right\rangle$ ***are*** the $\left\langle c_i^2 \right\rangle$. But we need to evaluate $\left\langle \underline{c}\,\underline{\delta f}^T \right\rangle$ in order to finish the job…

Let us evaluate $\left\langle \underline{c}\,\underline{\delta f}^T \right\rangle$ for the case where $\boxed{\underline{\delta f} = \underline{D}\,\underline{\delta m}}$.

Note that this is a ***linear*** relationship, with $\underline{D}$ being a $P \times P$ square matrix.

Then: $\underline{\delta f} = \underline{D}\,\underline{\delta m} \;\Rightarrow\; \left(\underline{f} - \underline{\hat{f}}\right) = \underline{D}\left(\underline{m} - \underline{\hat{m}}\right)$

Or: $\underline{f} = \underline{D}\,\underline{m} - \left(\underline{D}\,\underline{\hat{m}} - \underline{\hat{f}}\right) = \underline{D}\,\underline{m} - \left(\underline{D}\,\underline{\hat{m}} - \underline{\hat{m}}\right) = \underline{D}\,\underline{m} - \left(\underline{D} - \underline{\mathbb{1}}\right)\underline{\hat{m}}$

Now: $\underline{c} \equiv \underline{f} - \underline{m} = \left(\underline{f} - \underline{\hat{f}}\right) - \left(\underline{m} - \underline{\hat{m}}\right) = \underline{\delta f} - \underline{\delta m}$.

$$\therefore \quad \left\langle \underline{c}\,\underline{\delta f}^T \right\rangle = \left\langle \left(\underline{\delta f} - \underline{\delta m}\right)\underline{\delta f}^T \right\rangle = \left\langle \underline{\delta f}\,\underline{\delta f}^T \right\rangle - \left\langle \underline{\delta m}\,\underline{\delta f}^T \right\rangle = \underline{V}_{\underline{f}} - \left\langle \underline{\delta m}\,\underline{\delta f}^T \right\rangle$$

But from $\underline{\delta f} = \underline{D}\,\underline{\delta m}$ we get: $\underline{D} = \dfrac{\partial\left(\underline{\delta m}\right)}{\partial\left(\underline{\delta f}\right)}$, and from: $\underline{f} = \underline{D}\,\underline{m} - \left(\underline{D} - \underline{\mathbb{1}}\right)\underline{\hat{m}}$ we get: $\underline{D} = \dfrac{\partial \underline{f}}{\partial \underline{m}}$

$$\therefore \quad \left\langle \underline{c}\,\underline{\delta f}^T \right\rangle = \underline{V}_{\underline{f}} - \left\langle \underline{\delta m}\,\underline{\delta f}^T \right\rangle = \underline{V}_{\underline{f}} - \left\langle \underline{\delta m}\,\underline{\delta m}^T \underline{D}^T \right\rangle = \underline{V}_{\underline{f}} - \left\langle \underline{\delta m}\,\underline{\delta m}^T \right\rangle \underline{D}^T = \underline{V}_{\underline{f}} - \underline{V}_{\underline{m}}\underline{D}^T$$

Or: $\left\langle \underline{c}\,\underline{\delta f}^T \right\rangle = \underline{V}_{\underline{f}} - \left(\underline{D}\,\underline{V}_{\underline{m}}\right)^T$, since $\underline{V}_{\underline{m}} = \underline{V}_{\underline{m}}^T$ is a ***symmetric*** matrix.

Thus: $$\left\langle \underline{c}\,\underline{\delta f}^T \right\rangle = \underline{V}_{\underline{f}} - \left(\underline{D}\,\underline{V}_{\underline{m}}\right)^T = \underline{V}_{\underline{f}} - \left(\frac{\partial \underline{f}}{\partial \underline{m}}\underline{V}_{\underline{m}}\right)^T$$

But we earlier derived: $\dfrac{\partial \underline{f}}{\partial \underline{m}} = \mathbb{1} + \underline{V}_{\underline{m}} \underline{B} \underline{H}^{-1} \dfrac{\partial \underline{r}}{\partial \underline{m}} = \mathbb{1} - \underline{V}_{\underline{m}} \left(\underline{B} \underline{H}^{-1} \underline{B}^T \right)$ and: $\underline{V}_{\underline{f}} = \underline{V}_{\underline{m}} - \left(\underline{V}_{\underline{m}} \underline{B} \right) \underline{H}^{-1} \left(\underline{V}_{\underline{m}} \underline{B} \right)^T$

$$\therefore \quad \left\langle \underline{c}\, \underline{\delta f}^T \right\rangle = \underline{V}_{\underline{f}} - \left(\frac{\partial \underline{f}}{\partial \underline{m}} \underline{V}_{\underline{m}} \right)^T$$

$$= \underline{V}_{\underline{m}} - \left(\underline{V}_{\underline{m}} \underline{B} \right) \underline{H}^{-1} \left(\underline{V}_{\underline{m}} \underline{B} \right)^T - \left(\left[\mathbb{1} - \underline{V}_{\underline{m}} \left(\underline{B} \underline{H}^{-1} \underline{B}^T \right) \right] \underline{V}_{\underline{m}} \right)^T$$

$$= \cancel{\underline{V}_{\underline{m}}} - \cancel{\left(\underline{V}_{\underline{m}} \underline{B} \right) \underline{H}^{-1} \left(\underline{V}_{\underline{m}} \underline{B} \right)^T} - \cancel{\underline{V}_{\underline{m}}} + \cancel{\left(\underline{V}_{\underline{m}} \left(\underline{B} \underline{H}^{-1} \underline{B}^T \right) \underline{V}_{\underline{m}} \right)^T} = 0$$

Thus for the ***linear*** case where $\underline{\delta f} = \underline{D} \underline{\delta m}$: $\quad \underline{V}_{\underline{c}} \equiv \left\langle \underline{c}\, \underline{c}^T \right\rangle = \underline{V}_{\underline{m}} - \underline{V}_{\underline{f}} + 2 \left\langle \underline{c}\, \underline{\delta f}^T \right\rangle = \underline{V}_{\underline{m}} - \underline{V}_{\underline{f}}$

$$\therefore \quad p_i \equiv \frac{c_i}{\sqrt{\left\langle c_i^2 \right\rangle}} = \frac{f_i - m_i}{\sqrt{\left\langle \left(f_i - m_i \right)^2 \right\rangle}} \quad \Rightarrow \quad p_i \equiv \frac{c_i}{\sqrt{\left(\underline{V}_{\underline{c}} \right)_{ii}}} = \frac{f_i - m_i}{\sqrt{\left(\underline{V}_{\underline{m}} - \underline{V}_{\underline{f}} \right)_{ii}}} \quad \text{or:} \quad \boxed{p_i = \frac{f_i - m_i}{\sqrt{\sigma_{m_i}^2 - \sigma_{f_i}^2}}}$$

Naively, one might expect $\left\langle c_i^2 \right\rangle = \sigma_{f_i - m_i}^2 = \sigma_{m_i}^2 + \sigma_{f_i}^2$, but this ***ignores/neglects*** the ***correlation*** between m_i and f_i.

Since $\underline{V}_{\underline{f}} = \underline{V}_{\underline{m}} - \left(\underline{V}_{\underline{m}} \underline{B} \right) \underline{H}^{-1} \left(\underline{V}_{\underline{m}} \underline{B} \right)^T$, then $\sigma_{m_i}^2 > \sigma_{f_i}^2$ and thus we ***won't*** get into $\sqrt{\ }$ trouble in calculating the p_i "pulls".

Examples of LSQ fit "pulls" are shown in the figures below for a "Toy" Monte Carlo program that carries out LSQ fits to branching ratios of neutral and charged charmed D mesons, from a paper by Werner M. Sun, "Simultaneous least-squares treatment of statistical and systematic uncertainties", Nucl. Inst. Meth. Phys. Res. A 556 325-330 (2006).

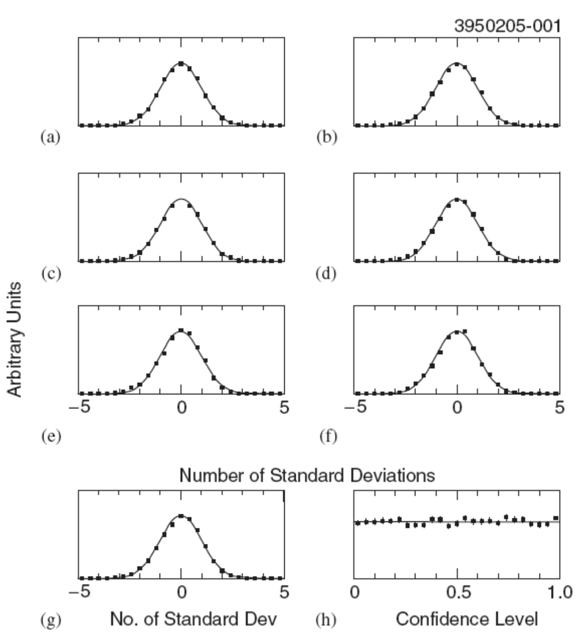

Fig. 1. Toy MC fit pull distributions for $\mathcal{N}^{00}$ (a), $\mathcal{B}(D^0 \to K^-\pi^+)$ (b), $\mathcal{B}(D^0 \to K^-\pi^+\pi^0)$ (c), $\mathcal{B}(D^0 \to K^-\pi^+\pi^-\pi^+)$ (d), $\mathcal{N}^{+-}$ (e), $\mathcal{B}(D^+ \to K^-\pi^+\pi^+)$ (f), and $\mathcal{B}(D^+ \to K_S^0\pi^+)$ (g), overlaid with Gaussian curves with zero mean and unit width. The fit confidence level distribution (*h*) is overlaid with a line with zero slope.

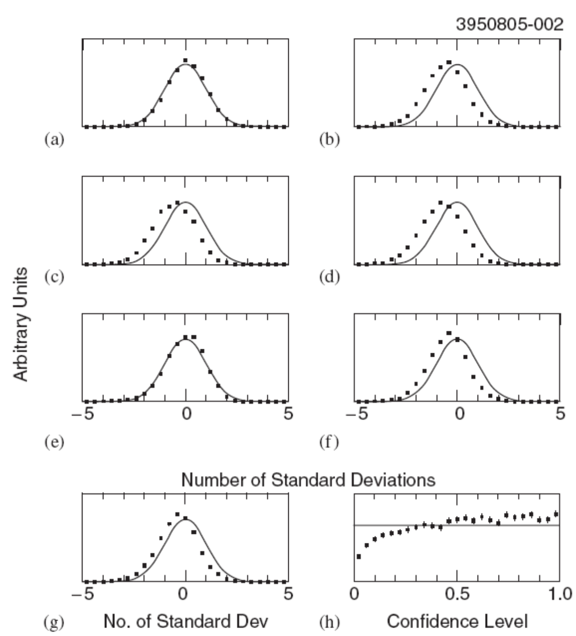

Fig. 2. Toy MC fit pull distributions, with $\mathbf{V_c}$ calculated using $\mathbf{c}$ instead of $\tilde{\mathbf{c}}$, for $\mathcal{N}^{00}$ (a), $\mathcal{B}(D^0 \to K^-\pi^+)$ (b), $\mathcal{B}(D^0 \to K^-\pi^+\pi^0)$ (c), $\mathcal{B}(D^0 \to K^-\pi^+\pi^-\pi^+)$ (d), $\mathcal{N}^{+-}$ (e), $\mathcal{B}(D^+ \to K^-\pi^+\pi^+)$ (f), and $\mathcal{B}(D^+ \to K_S^0\pi^+)$ (g), overlaid with Gaussian curves with zero mean and unit width. The fit confidence level distribution (*h*) is overlaid with a line with zero slope.