Glen Cowan
RHUL Physics
23 September 2019

Parameter Estimation with Constraints

This note summarizes the discussions between GDC and Olaf Behnke on parameter estimation with constraints. It provides details and derivations of the material that appears in condensed form in Sec. 40.2.4 of the PDG Review of Particle Properties [1].

In some applications one is interested in using a set of measured quantities $\mathbf{y} = (y_1, \ldots, y_N)$ to estimate a set of parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_M)$ subject to a number of constraints. For example, one may have measured coordinates from two tracks, and one wishes to estimate their momentum vectors subject to the constraint that the tracks have a common vertex. The parameters can also include momenta of undetected particles such as neutrinos, as long as the constraints from conservation of energy and momentum and from known masses of particles involved in the reaction chain provide enough information for these quantities to be inferred.

A set of K constraints can be given in the form of equations

$$c_k(\boldsymbol{\theta}) = 0 \;, \quad k = 1, \ldots, K \;. \tag{1}$$

In some problems it may be possible to define a new set of parameters $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_L)$ with $L = M - K$ such that every point in $\boldsymbol{\eta}$-space automatically satisfies the constraints. If this is possible then the problem reduces to one of estimating $\boldsymbol{\eta}$ with, e.g., maximum likelihood or least squares and then transforming the estimators back into $\boldsymbol{\theta}$-space.

In many cases it may be difficult or impossible to find an appropriate transformation $\boldsymbol{\eta}(\boldsymbol{\theta})$. Suppose that the parameters are determined through minimizing an objective function such as $\chi^2(\boldsymbol{\theta})$ in the method of least squares. Here one may enforce the constraints by finding the stationary points of the *Lagrange function*

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\lambda}, \mathbf{y}) = \chi^2(\boldsymbol{\theta}, \mathbf{y}) + \sum_{k=1}^{K} \lambda_k c_k(\boldsymbol{\theta}) \tag{2}$$

with respect to both the parameters $\boldsymbol{\theta}$ and a set of *Lagrange multipliers* $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_K)$. Combining the parameters and Lagrange multipliers into an $(M + K)$-component vector $\boldsymbol{\gamma} = (\theta_1, \ldots, \theta_M, \lambda_1, \ldots, \lambda_K)$, the solutions for $\boldsymbol{\gamma}$, i.e., the estimators $\hat{\boldsymbol{\gamma}}$, are found (e.g., numerically) from the system of equations

$$F_i(\boldsymbol{\gamma}, \mathbf{y}) \equiv \frac{\partial \mathcal{L}}{\partial \gamma_i} = 0 \;, \quad i = 1, \ldots, M + K \;. \tag{3}$$

To obtain the covariance matrix of the estimated parameters one can find solutions $\tilde{\boldsymbol{\gamma}}$ corresponding to the expectation values of the data $\langle \mathbf{y} \rangle$ and expand $F_i(\hat{\boldsymbol{\gamma}}, \mathbf{y})$ to first order about these values. This gives (see, e.g., Sec. 11.6 of Ref. [2]) linearized approximations for the estimators, $\hat{\boldsymbol{\gamma}}(\mathbf{y}) \approx \tilde{\boldsymbol{\gamma}} + C(\mathbf{y} - \langle \mathbf{y} \rangle)$, where the matrix $C = -A^{-1}B$, and A and B are given by

$$A_{ij} = \left[\frac{\partial F_i}{\partial \gamma_j} \right]_{\tilde{\boldsymbol{\gamma}}, \langle \mathbf{y} \rangle} \quad \text{and} \quad B_{ij} = \left[\frac{\partial F_i}{\partial y_j} \right]_{\tilde{\boldsymbol{\gamma}}, \langle \mathbf{y} \rangle} \;. \tag{4}$$

In practice the values $\langle \mathbf{y} \rangle$ and corresponding solutions $\tilde{\boldsymbol{\gamma}}$ are estimated using the data from the actual measurement. Using this approximation for $\hat{\boldsymbol{\gamma}}(\mathbf{y})$, one can find the covariance matrix $U_{ij} = \mathrm{cov}[\hat{\gamma}_i, \hat{\gamma}_j]$ of the the estimators for the γ_i in terms of that of the data $V_{ij} = \mathrm{cov}[y_i, y_j]$ using error propagation ,

$$U = CVC^T \ . \tag{5}$$

The upper-left $M \times M$ block of the matrix U gives the covariance matrix for the estimated parameters $\mathrm{cov}[\hat{\theta}_i, \hat{\theta}_j]$. If the parameters are estimated using the method of least squares, then the number of degrees of freedom for the distribution of the minimized χ^2 is increased by the number of constraints, i.e., it becomes $N - M + K$. Further details can be found in, e.g., Ch. 7 of Ref. [3].

References

[1] R.L. Workman et al. (Particle Data Group), Prog. Theor. Exp. Phys. 2022, 083C01 (2022) and 2023 update; `pdg.lbl.gov`.

[2] G. Cowan, *Statistical Data Analysis*, (Oxford University Press, Oxford, 1998).

[3] Olaf Behnke et al., *Data Analysis in High Energy Physics: A Practical Guide to Statistical Methods*, Wiley, 2013.

[4] V. Blobel and E. Lohrmann, *Statistische und numerische Methoden der Datenanalyse*, Teubner Verlag (1998); e-Buch: `www.desy.de/~blobel/eBuch.pdf` (2012).

[5] Wikipedia page for *block matrix*, `en.wikipedia.org/wiki/Block_matrix`.

Appendix A: Derivation of covariance matrix

Starting from the equations $F_i(\boldsymbol{\gamma}, \mathbf{y}) = 0$, $i = 1, \ldots, K + M$, consider two solutions: $\hat{\boldsymbol{\gamma}}$ corresponding to data $\mathbf{y}$ and $\tilde{\boldsymbol{\gamma}}$ corresponding to $\langle \mathbf{y} \rangle$. Expanding $F_i(\hat{\boldsymbol{\gamma}}, \mathbf{y})$ to first order in $\hat{\boldsymbol{\gamma}}$ and $\mathbf{y}$ about $\tilde{\boldsymbol{\gamma}}$ and $\langle \mathbf{y} \rangle$ gives

$$F_i(\mathbf{y}, \hat{\boldsymbol{\gamma}}) \approx F_i(\langle \mathbf{y} \rangle, \tilde{\boldsymbol{\gamma}}) + \sum_{j=1}^{M+K} \left[\frac{\partial F_i}{\partial \gamma_j} \right]_{\langle \mathbf{y} \rangle, \tilde{\boldsymbol{\gamma}}} (\hat{\gamma}_j - \tilde{\gamma}_j) \ + \ \sum_{j=1}^{N} \left[\frac{\partial F_i}{\partial y_j} \right]_{\langle \mathbf{y} \rangle, \tilde{\boldsymbol{\gamma}}} (y_j - \langle y_j \rangle) \ . \tag{6}$$

The terms $F_i(\mathbf{y}, \hat{\boldsymbol{\gamma}})$ and $F_i(\langle \mathbf{y} \rangle, \tilde{\boldsymbol{\gamma}})$ are both zero because both pairs of arguments are assumed to be solutions to $F_i = 0$. Dropping these terms, the equation can be rewritten in matrix form $\hat{\boldsymbol{\gamma}} \approx \tilde{\boldsymbol{\gamma}} + C(\mathbf{y} - \langle \mathbf{y} \rangle)$, where $C = -A^{-1}B$ and the definitions of the matrices A and B are as given in Eq. (4).

Appendix B: Comments on notation

In the notation used above in the example of a kinematic fit, the parameters $\boldsymbol{\theta}$ are meant to include all of the quantities to be estimated, including true momenta of particles for which there are measured values as well as those like neutrinos that are not measured. The vector $\mathbf{y}$ represents all of the measured quantities, and the χ^2 term would usually have the form

$$\chi^2(\boldsymbol{\theta}, \mathbf{y}) = (\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\theta}))^T V^{-1} (\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\theta})) \ . \tag{7}$$

Here $\boldsymbol{\mu}$ means the expectation values of the measurements, i.e.,

$$E[y_i] = \mu_i(\boldsymbol{\theta}) \ . \tag{8}$$

This covers the case where a component of $\boldsymbol{\theta}$ includes a parameter for which there is no direct measurement, such as a neutrino momentum. The various μ_i would then not depend directly on such a parameter, but because of the constraint terms $c_k(\boldsymbol{\theta})$, one nevertheless obtains estimators for them. That is, if θ_j is something like a neutrino momentum, then $\partial \chi^2 / \partial \theta_j = 0$, but because this θ_j enters into the c_k terms, the Lagrange function depends on θ_j, allowing one to estimate it.

Appendix C: Comparison to Blobel and Lohrmann

A similar treatment of constrained fits is given in the German-language monograph of Blobel and Lohrmann [4], Ch. 7. In their notation, from Eq. (7.114) the covariance matrix of the estimated parameters corresponds to C_{22} , which by Eq. (7.112) is

$$C_{22} = W_A^{-1} \ . \tag{9}$$

Then according to Eq. (7.111) we have $W_A = (A^T W_B A)$, and therefore

$$\begin{aligned} C_{22} &= (A^T W_B A)^{-1} \\ &= A^{-1} W_B^{-1} (A^T)^{-1} \\ &= A^{-1} W_B^{-1} (A^{-1})^T \ , \end{aligned} \tag{10}$$

where the last line follows from $(A^T)^{-1} = (A^{-1})^T$. Then using Eq. (7.110), $W_B = (B W^{-1} B^T)^{-1}$, we have

$$\begin{aligned} C_{22} &= A^{-1} B W^{-1} B^T (A^{-1})^T \\ &= A^{-1} B V (A^{-1} B)^T \ , \end{aligned} \tag{11}$$

where $V = W^{-1}$ is the covariance matrix of the measurements, $V_{ij} = \mathrm{cov}[y_i, y_j]$. Then in GDC's notation we have $U = C_{22}$ and $C = -A^{-1} B$, and therefore

$$U = C V C^T \ , \tag{12}$$

which confirms the desired result.

Appendix D: Simplified formula for covariance matrix

In 2017 it was pointed out by Olaf Behnke that Eq. (5) for the covariance matrix U of the constrained estimators can, with some restrictions, be expressed in a simpler form, namely, U is the upper-left $M \times M$ submatrix of $2A^{-1}$, where the $(M+K) \times (M+K)$ matrix A is defined by the first of Eqs. (4). This follows from Eq. (7.99) in the book by Blobel and Lohrmann [4]. Below a rederivation along these lines is given using the notation of this note.

First, suppose that the measurements $\mathbf{y} = (y_1, \ldots, y_N)^T$ have expectation values that can be expressed in terms of the parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_M)^T$ as a linear relation,

$$E[\mathbf{y}] = R\boldsymbol{\theta} \; . \tag{13}$$

Here $\mathbf{y}$ and $\boldsymbol{\theta}$ are column vectors and R is an $N \times M$ matrix.

In addition, suppose that the constraints are given by linear functions of the parameters. That is, the functions $c_k(\boldsymbol{\theta})$ with $k = 1, \ldots, K$ can be expressed as a first-order Taylor series about an arbitrary fixed point $\tilde{\boldsymbol{\theta}}$ as

$$c_k(\boldsymbol{\theta}) = c_k(\tilde{\boldsymbol{\theta}}) + \sum_{i=1}^{M} \frac{\partial c_k}{\partial \theta_i} (\theta_i - \tilde{\theta}_i) \; , \tag{14}$$

and that the derivatives

$$\frac{\partial c_k}{\partial \theta_i} \equiv a_{ki} \tag{15}$$

are constants (independent of $\boldsymbol{\theta}$), which we will write using the $K \times M$ matrix a. The constraint functions can thus be written as the vector equation

$$\mathbf{c}(\boldsymbol{\theta}) = \mathbf{c}(\tilde{\boldsymbol{\theta}}) + a(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) \; . \tag{16}$$

The derivation below requires both a linear least-squares problem (expectation values of measurements linear in the parameters) and linear constraints. If these restrictions are not met then the final result for the covariance matrix is at best an approximation.

The solution to the unconstrained problem is found by minimizing

$$\chi^2(\boldsymbol{\theta}) = (\mathbf{y} - R\boldsymbol{\theta})^T V^{-1} (\mathbf{y} - R\boldsymbol{\theta}) \; . \tag{17}$$

The resulting estimators will be denoted with primes and are found to be

$$\boldsymbol{\theta}' = (R^T V^{-1} R)^{-1} R^T V^{-1} \mathbf{y} = W^{-1} R^T V^{-1} y \; , \tag{18}$$

where $W = R^T V^{-1} R$ can also be related to the second derivative of χ^2 as

$$W_{ij} = \frac{1}{2} \frac{\partial^2 \chi^2}{\partial \theta_i \partial \theta_j} \bigg|_{\boldsymbol{\theta}'} \; . \tag{19}$$

One finds that the covariance matrix of the unconstrained estimators $U'_{ij} = \mathrm{cov}[\theta'_i, \theta'_j]$ is given by $U' = W^{-1}$.

To find the least-squares estimators subject to the constraints $\mathbf{c}(\boldsymbol{\theta}) = 0$, we seek, as before, the minimum of the Lagrange function

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \chi^2(\boldsymbol{\theta}) + \sum_{k=1}^{K} \lambda_k c_k(\boldsymbol{\theta}) \tag{20}$$

with respect to $\boldsymbol{\theta}$ as well as the Lagrange multipliers $\boldsymbol{\lambda}$. Defining as above the $(M+K)$-component column vector $\boldsymbol{\gamma} = (\boldsymbol{\theta}, \boldsymbol{\lambda})$, the solutions can be found from

$$\mathbf{F}(\boldsymbol{\gamma}) = \nabla_{\boldsymbol{\gamma}} \mathcal{L} = 0 \; , \tag{21}$$

where $\mathbf{F} = (F_1, \ldots, F_{M+K})^T$ is column vector of $M + K$ functions and the operator $\nabla_{\boldsymbol{\gamma}}$ denotes the vector of derivatives with respect to the components of $\boldsymbol{\gamma}$.

The estimators $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}})$ satisfy $\mathbf{F}(\hat{\boldsymbol{\gamma}}) = 0$. To find the solutions we expand $\mathbf{F}$ about $\boldsymbol{\gamma}' = (\boldsymbol{\theta}', \boldsymbol{\lambda}')$, where $\boldsymbol{\theta}'$ is the solution to the unconstrained problem found above and $\boldsymbol{\lambda}'$ is an arbitrary set of values for the Lagrange multipliers, which in the final result cancel out. This gives

$$\mathbf{F}(\hat{\boldsymbol{\gamma}}) = \mathbf{F}(\boldsymbol{\gamma}') + A(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}') \; , \tag{22}$$

where the matrix A is defined as

$$A_{ij} = \left. \frac{\partial F_i}{\partial \gamma_j} \right|_{\boldsymbol{\gamma}'} \; . \tag{23}$$

The matrix A differs from that defined in Eq. (4) only in that the derivatives are evaluated at the unconstrained estimators rather than at the (mean) solutions to the constrained problem.

The components of $\mathbf{F}$ are

$$F_i(\boldsymbol{\gamma}) = \begin{cases} \frac{\partial \chi^2}{\partial \theta_i} + \sum_{k=1}^{K} \frac{\partial c_k(\boldsymbol{\theta})}{\partial \theta_i} \lambda_k & i = 1, \ldots, M \; , \\ c_{i-M}(\boldsymbol{\theta}) & i = M+1, \ldots, M+K. \end{cases} \tag{24}$$

Evaluating these at $\boldsymbol{\gamma}'$ gives

$$\mathbf{F}(\boldsymbol{\gamma}') = \begin{pmatrix} a^T \boldsymbol{\lambda}' \\ c(\boldsymbol{\theta}') \end{pmatrix} \; , \tag{25}$$

where the matrix a is defined in Eq. (15) (recall that for linear constraints that a is independent of $\boldsymbol{\theta}$ and that $\partial \chi^2 / \partial \theta_i = 0$ when evaluated at the unconstrained solution $\boldsymbol{\theta}'$).

For the derivatives evaluated at $\boldsymbol{\gamma}'$ we find

$$A_{ij} = \frac{\partial F_i}{\partial \gamma_j}\bigg|_{\gamma'} = \begin{cases} \frac{\partial^2 \chi^2}{\partial\theta_i \partial\theta_j}\Big|_{\boldsymbol{\theta}'} = 2W & i = 1, \ldots, M,\ j = 1, \ldots, M\ , \\ \frac{\partial c_{j-M}}{\partial \theta_i} = a_{ki} & i = 1, \ldots, M,\ j = M+1, \ldots, M+K,\ k = j - M\ , \\ \frac{\partial c_{i-M}}{\partial \theta_j} = a_{kj} & i = M+1, \ldots, M+K,\ j = 1, \ldots, M,\ k = i - M, \\ 0 & i = M+1, \ldots, M+K,\ j = M+1, \ldots, M+K. \end{cases} \tag{26}$$

Note that terms involving the second derivatives of $\mathbf{c}(\boldsymbol{\theta})$ do not appear as we are assuming linear constraints. The matrix A can thus be written

$$A = \begin{pmatrix} 2W & a^T \\ a & 0 \end{pmatrix}\ , \tag{27}$$

and Eq. (22) therefore becomes

$$\mathbf{F}(\hat{\gamma}) = \begin{pmatrix} a^T \boldsymbol{\lambda}' \\ c(\boldsymbol{\theta}') \end{pmatrix} + \begin{pmatrix} 2W & a^T \\ a & 0 \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}' \\ \hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}' \end{pmatrix} = 0\ . \tag{28}$$

In the first M rows of this system of equations, the terms $a^T \boldsymbol{\lambda}'$ cancel and one finds

$$2W(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}') + a^T \hat{\boldsymbol{\lambda}} = 0\ . \tag{29}$$

Rows $M + 1$ to $M + K$ give

$$\mathbf{c}(\boldsymbol{\theta}') + a(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}') = 0\ . \tag{30}$$

We also have Eq. (14) which can be used to express $\mathbf{c}(\boldsymbol{\theta}')$ as

$$\mathbf{c}(\boldsymbol{\theta}') = \mathbf{c}(\tilde{\boldsymbol{\theta}}) + a(\boldsymbol{\theta}' - \tilde{\boldsymbol{\theta}})\ , \tag{31}$$

where $\tilde{\boldsymbol{\theta}}$ is an arbitrary constant expansion point. The idea is now to use Eqs. (29), (30) and (31) to eliminate $\mathbf{c}(\boldsymbol{\theta}')$ and $\hat{\boldsymbol{\lambda}}$ and thus give a relation that can be solved for $\hat{\boldsymbol{\theta}}$ in terms of $\boldsymbol{\theta}'$ along with constant terms. Then by using the known relation between $\boldsymbol{\theta}'$ and the original data $\mathbf{y}$ together with error propagation we can find the covariance matrix of the constrained estimators $\hat{\boldsymbol{\theta}}$.

Equation (29) can be rewritten as (cf. Eq. (7.95) in [4])

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}' = -\frac{1}{2} W^{-1} a^T \hat{\boldsymbol{\lambda}}\ . \tag{32}$$

Substituting into Eq. (30) and solving for $\hat{\boldsymbol{\lambda}}$ gives (cf. Ref. [4], Eq. (7.96)),

$$\hat{\boldsymbol{\lambda}} = 2(aWa^T)^{-1} \mathbf{c}(\boldsymbol{\theta}')\ . \tag{33}$$

Now using Eq. (31) for $\mathbf{c}(\boldsymbol{\theta}')$ in Eq. (33) gives

$$\hat{\boldsymbol{\lambda}} = 2(aWa^T)^{-1}\left[\mathbf{c}(\tilde{\boldsymbol{\theta}}) + a(\boldsymbol{\theta}' - \tilde{\boldsymbol{\theta}})\right] . \tag{34}$$

Substituting this back into Eq. (29) and solving for $\hat{\boldsymbol{\theta}}$ gives (cf. Ref. [4], Eq. (7.98))

$$\hat{\boldsymbol{\theta}} = \left[I - W^{-1}a^T(aW^{-1}a^T)^{-1}a\right]\boldsymbol{\theta}' + \text{const.}, \tag{35}$$

where the constant term includes the fixed expansion point $\tilde{\boldsymbol{\theta}}$ as well as the matrices a and W. Here a is constant as it is independent of $\boldsymbol{\theta}$ for linear constraints and thus does not depend on where the derivatives of $\mathbf{c}(\boldsymbol{\theta})$ are evaluated. Furthermore $W = R^TV^{-1}R$ is independent of the data provided the expectation values of the data are linear in the parameters and is thus also a constant given that restriction.

Now to find the covariance of $\hat{\boldsymbol{\theta}}$ we recall that the covariance of the unconstrained estimators $\boldsymbol{\theta}'$ is $U' = W^{-1}$. Furthermore define $W_a = (aW^{-1}a^T)^{-1}$. Then using error propagation one has for $U_{ij} = \text{cov}[\hat{\theta}_i, \hat{\theta}_j]$

$$\begin{aligned} U &= \left(I - W^{-1}a^TW_aa\right)U'\left(I - W^{-1}a^TW_aa\right)^T \\ &= W^{-1} - 2W^{-1}a^TW_aaW^{-1} + W^{-1}a^TW_aaW^{-1}a^TW_aaW^{-1} . \end{aligned} \tag{36}$$

Using $aW^{-1}a^T = W_a^{-1}$ in the last line above gives the final result

$$\begin{aligned} U &= W^{-1} - W^{-1}a^TW_aaW^{-1} \\ &= W^{-1} - W^{-1}a^T(aW^{-1}a^T)^{-1}aW^{-1} . \end{aligned} \tag{37}$$

This formula for the covariance of the constrained estimators is one way of expressing the final result of this section. It can also be directly related to the matrix A as given in Eq. (27). The inverse of a 2×2 block matrix can be written (see, e.g., Ref. [5]),

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{pmatrix} . \tag{38}$$

For a matrix of the form of A in Eq. (27) by substituting $A \to 2W$, $B \to a^T$, $C \to a$ and $D \to 0$ one finds

$$A^{-1} = \begin{pmatrix} 2W & a^T \\ a & 0 \end{pmatrix}^{-1} = \frac{1}{2}\begin{pmatrix} W^{-1} - W^{-1}a^T(aW^{-1}a^T)^{-1}aW^{-1} & 2W^{-1}a^T(aW^{-1}a^T)^{-1} \\ 2(aW^{-1}a^T)^{-1}aW^{-1} & -4(aW^{-1}a^T)^{-1} \end{pmatrix} . \tag{39}$$

A somewhat more convenient form is found using $H = 2W$, i.e.,

$$H_{ij} = \left.\frac{\partial^2\chi^2}{\partial\theta_i\partial\theta_j}\right|_{\boldsymbol{\theta}'} \tag{40}$$

is the Hessian matrix of χ^2 evaluated at its unconstrained minimum. Equation (39) becomes

$$A^{-1} = \begin{pmatrix} H & a^T \\ a & 0 \end{pmatrix}^{-1} = \begin{pmatrix} H^{-1} - H^{-1}a^T(aH^{-1}a^T)^{-1}aH^{-1} & H^{-1}a^T(aH^{-1}a^T)^{-1} \\ (aH^{-1}a^T)^{-1}aH^{-1} & -(aH^{-1}a^T)^{-1} \end{pmatrix} . \tag{41}$$

By comparing the expression for the covariance matrix U found in Eq. (37) to either of Eqs. (39) or (41) one sees that U is given by the upper-left $(M \times M)$ submatrix of $2A^{-1}$, thus proving the result first spotted by Olaf.

This result comes with some restrictions, namely, the constraints must be linear and the expectation values of the measurements must be linear in the parameters. One can see how these conditions are relevant as the matrix A used to compute $U = [2A^{-1}]_{M\times M}$ is defined as $\partial F_i/\partial\theta_j$ evaluated at the *unconstrained* estimators, whereas the corresponding matrix defined in the first of Eqs. (4) is evaluated at solutions to the *constrained* problem. Provided both linearity requirements hold, the matrices a and W that enter into A are independent of the parameters and thus it does not matter whether one evaluates at $\hat{\boldsymbol{\theta}}$ or $\boldsymbol{\theta}'$. If the linearity conditions do not hold then the equivalence between $U = 2A^{-1}$ found here and $U = CVC^T$ with $C = -A^{-1}B$ from Eq.(5) is only approximate.