DEGREES OF FREEDOM DETERMINATION IN ACCELERATOR PHYSICS OPTIMISATION PROBLEMS SUBJECT TO GLOBAL CONSTRAINTS

A.-S. Müller, CERN, Geneva, Switzerland

Abstract

The determination of the number of degrees of freedom of a system is a common problem in physics. It is straightforward for the case of unbounded parameters. For global constraints cutting into the parameter space however, the determination is difficult. This problem occurs, for example, in the matching of beam lines where global constraints like aperture or maximum bend angle have to be respected. It is also relevant for fits of complex models to experimental data, where external information (e.g. positivity, maximum energy loss or measured momentum spread) is included in the form of constraints. This paper proposes a method to extract the effective number of degrees of freedom for a given system. Examples are discussed to illustrate the method.

1 INTRODUCTION

The counting of degrees of freedom ("DoF") is a problem encountered frequently throughout various regions of physics though the meaning of the term "degrees of freedom" may vary from case to case. In classical mechanics, the DoF of a system denote its possible independent directions of motion. In fitting problems, the number of DoF is defined by the number of observations and free parameters and is used as a quality measure for the fit; a χ^2/DoF $\cong 1$ usually denotes a consistent fit with well adapted measurement errors and an appropriate model function. If the χ^2/DoF $\ncong 1$ for a well known model function, its square root yields an estimator for the true uncertainty. In the matching of a beam line, the number of DoF means the number of independent combinations of quadrupole magnets available to achieve a certain set of Twiss parameters at a given place. The last two examples belong to the wide field of optimisation problems: for both, the task consists in finding the best sets of parameters for the given situation. Usually, the parameters involved in the optimisation are unbounded and it is straightforward to determine the number of DoF. For global constraints cutting into the parameter space by bounding the parameter value to a specified region with a given precision, the procedure of counting is no longer well defined. In the matching of beam lines, such global constraints can be imposed by the vacuum chamber's aperture or a maximum bend angle. For fits of complex models to experimental data, constraints are used to include external information (e.g. positivity, maximum energy loss or measured momentum spread).

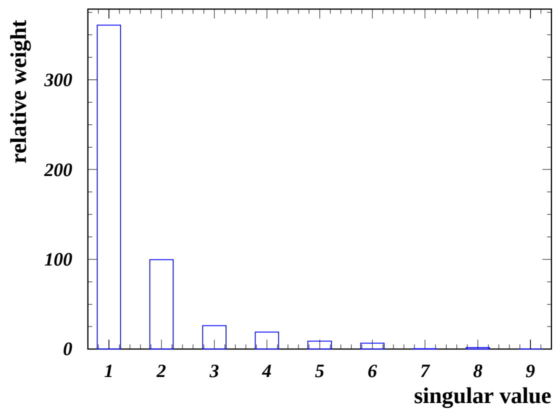

Figure 1: Spectrum of singular values (components of the diagonal matrix) for the response matrix of the extented T9 beam line of the CERN PS East Hall facility.

2 BEAM LINE MATCHING

In the matching of beam lines, the parameters to be adjusted are typically the normalised gradients of the n available quadrupole magnets. Positions of the elements or gradients of higher order elements are also possible, but for simplicity we will confine ourselves to the example of linear optics. Instead of observables, m functions at a point of interest in the line (e.g. focal point, stripping foil) are used that depend on all parameters or combinations of them (e.g. group of magnets used for dispersion correction). The optical parameters constitute a vector of dimension $m = 8$: $\vec{f}(k_1, \ldots, k_n) = \{\alpha_h, \alpha_v, \beta_h, \beta_v, D_h, D_v, D'_h, D'_v\}$.

The change of the function vector with respect to the parameter (i.e. the gradients) can be described as $\Delta\vec{f} = A\,\Delta\vec{k}$, where the elements of the response matrix A are $A_{ij} = \frac{\partial f_i}{\partial k_j}$. An evaluation of A is done by Monte Carlo: all free parameters (k) are varied and MAD [1] is used to calculate f_i at the point of interest. Only solutions are permitted that respect the implemented global constraints. The correlations are given by $\rho_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}}$ with $C_{ij} = \langle v_i v_j\rangle - \langle v_i\rangle\langle v_j\rangle$ and $\vec{v} = \{k_1, \ldots, k_n, f_1, \ldots, f_m\}$. The correlation matrix is related to A by

$$\rho_{ij} = \frac{\partial f_i}{\partial k_j}\sqrt{\frac{C_{jj}}{C_{ii}}} = A_{ij}\sqrt{\frac{C_{jj}}{C_{ii}}}. \qquad (1)$$

To extract the effective number of degrees of freedom of the problem, a singular value decomposition (SVD) is applied to the response matrix A, normalised with the optimal $\bar{k}/\bar{f}$ in order to get rid of absolute scales. A measure for the effective number of DoF is $W = (\sum_{i=1}^{n} s_i)^2 / \sum_{i=1}^{n} s_i^2$ where s_i are the singular values. The example in Fig. 1 already indicates that although there are nine quadrupoles present, the problem possesses only $W = 1.9$ effective degrees of freedom. An other example for a drastic reduction of the DoF can be found in [2].

3 MULTI PARAMETER FIT

Very often, physical quantities of interest can only be determined in an indirect way. In this case, some observable y_i, depending on the parameter or parameters of interest ($\vec{a}$), is measured as a function of one of the parameters it depends on (x_i). A model $g(x_i, \vec{a})$ describing this dependence is then fitted to the set of individual observations $\{x_i, y_i\}$ which means finding the set of estimators that brings the model as close as possible to the average of the observations. The "distance" between model and observations is defined by a weight function. The best set of estimators for the true parameters is obtained for the minimum of this weight function. A commonly used weight function is χ^2 which for the general case, where both measurements y_i and the model function g can be vector quantities, is given by

$$\chi^2 \;=\; \sum_i \left(y_i - g(x_i, \vec{a})\right)^T C_i^{-1} \left(y_i - g(x_i, \vec{a})\right) \qquad (2)$$

with covariance matrix C.
The distribution of χ^2 follows the probability density function (PDF)

$$f(\chi^2) \;=\; \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} \,(\chi^2)^{\frac{n-2}{2}}\, e^{-\frac{1}{2}(\chi^2)} \qquad (3)$$

where n is the number of degrees of freedom in the optimisation problem. For a completely unconstrained fit, where all fit parameters can vary without limits, n corresponds to the number of individual observations minus the number of fit parameters (or dimension of $\vec{a}$): $n = N - p$. For optimisation problems where parameters are fixed, the number of DoF n increases by one for each fixed parameter. Since the Gamma function is defined for all real numbers $x > 0$, the number of degrees of freedom need not necessarily be integer from a technical point of view. The PDF of the χ^2 distribution is therefore also defined for non-integer degrees of freedom. This becomes of practical use, when external knowledge has to be included in the fit in the form of parameter constraints with a certain precision. For example, if one of the parameters is known to be positive or has been measured independently, a term can be added to the χ^2 function to account for this. The exact form of the constraint will vary with the special case. For Gaussian uncertainties of the independently determined parameter, a constraint of the form $((a_i - a_{i\mathrm{nom}})/S_{a_i})^2$ is the natural choice, where $a_{i\mathrm{nom}}$ is the value the parameter is constrained to, and S_{a_i} is its uncertainty. A variation of this "uncertainty" modifies the impact of the constraint on the χ^2 function and can therefore be used to define the strength of the constraint $K_\sigma = S/\sigma_0$, with σ_0 being the uncertainty of the unconstrained parameter derived from the fit. To summarise: depending on K_σ the parameter will either be free ($N - p$) or fixed ($N - p + 1$) or something in between. In the latter case it is not obvious to count the number of DoF and we will have to introduce non-integer DoF to account for the constrained parameter variation. For this we can use the interesting properties of the χ^2 distribution, that both mean and variance reflect the number of degrees of freedom:

$$\langle\chi^2\rangle \;=\; n \quad \text{and} \quad V(\chi^2) \;=\; 2n.$$

With the PDF being defined for all real numbers, the mean can therefore be used to determine the effective number of DoF of a constrained problem. For m parameters constraint, the extended $\langle\chi^2\rangle$ (usually the one for the best set of estimators $\langle\chi^2_{\min}\rangle$) is

$$\langle\chi^2_{\min}\rangle_E \;=\; N - p \;+\; \sum_{i=1}^{m\le p} \frac{1}{1 + K_{i\sigma}^2}. \qquad (4)$$

As an illustration, the expression for the evolution of the average χ^2 with the constraint of one parameter is derived for the case of a fluctuation around a constant ($p = 1$). Without loss of generality we can assume this constant to be zero. The χ^2 function for this case is given by

$$\chi^2 \;=\; \frac{a^2}{S^2} \;+\; \sum_{i=1}^{N}\left(\frac{a - x_i}{\sigma_i}\right)^2 \qquad (5)$$

where σ_i is the error of the individual observation x_i, a the single component of the vector $\vec{a}$ in Eq.(2) and S is its constraint. From the first derivative of χ^2 with respect to parameter a, the estimator of a at the minimum χ^2 is

$$a_0 \;=\; \frac{1}{w}\sum_{i=1}^{N}\frac{x_i}{\sigma_i^2} \quad \text{with} \quad w \;=\; \frac{1}{S^2} + \sum_{i=1}^{N}\frac{1}{\sigma_i^2}. \qquad (6)$$

Using this, the expression for χ^2 at its minimum becomes

$$\chi^2_{\min} \;=\; \sum_{i=1}^{N}\left(\frac{x_i^2}{\sigma_i^2} - \frac{a_0 x_i}{\sigma_i^2}\right). \qquad (7)$$

With $\langle x_i\rangle = 0$ and $\langle x_i^2\rangle = \sigma_i^2$ the expectation value for $\chi^2_{\min}$ is

$$\langle\chi^2_{\min}\rangle = N - \frac{1}{w}\langle\sum_{i=1}^{N}\frac{x_i}{\sigma_i^2}\sum_{j=1}^{N}\frac{x_j}{\sigma_j^2}\rangle = N - \frac{1}{w}\sum_{i=1}^{N}\frac{1}{\sigma_i^2}.$$

The sum over the inverse squares of the observation uncertainties is exactly the error of the parameter of a weighted mean and is therefore interpreted as the uncertainty of the unconstrained parameter in the fit:

$$\left(\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\right)^{-1} \;:=\; \sigma_0^2$$

The mean of the extended χ^2 function for this simple example is finally

$$\langle\chi^2_{\min}\rangle_E \;=\; N - 1 \;+\; \frac{1}{1 + (S/\sigma_0)^2}. \qquad (8)$$

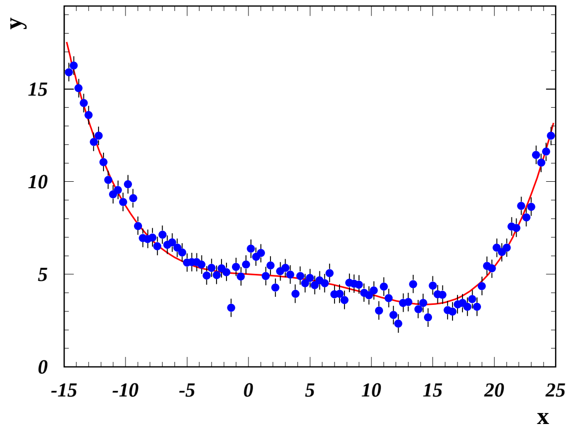

Figure 2: Model function used for numerical cross-check of relation Eq.(4). The curve represents Eq.(9), and the 100 points are Gaussian variations around it, simulating the observations $\{x_i, y_i\}$.

With $K_\sigma = S/\sigma_0$ this is precisely Eq.(4) for one parameter.

A numerical cross-check of the validity of the relation Eq.(4) is done with a polynomial as model function:

$$g(x, \vec{a}) = \sum_{k=1}^{5} a_k x^{k-1} \tag{9}$$

with $a_1 = 5$, $a_2 = -0.04$, $a_3 = 0.003$, $a_4 = -0.002$ and $a_5 = 0.0001$. This function is shown in Fig. 2, represented by a curve. The 100 points in Fig. 2 are Gaussian variations around the model function, simulating the individual observations $\{x_i, y_i\}$. A series of 10^5 fits with different random variations to simulate the observations is done with different strengths of a constraint of parameter a_1. From the resulting χ^2_{min} the mean is calculated to extract a number of degrees of freedom. This result is displayed as a function of the strength K_σ of the constraint of parameter a_1 in Fig. 4. To clearly show the functional dependence, $\ln K_\sigma$ is plotted. The curve is a fit of the extended $\langle\chi^2_{\text{min}}\rangle_E$ defined in Eq.(4). An example of an analysis with external knowledge included as constraints in the described way can be found in [3].

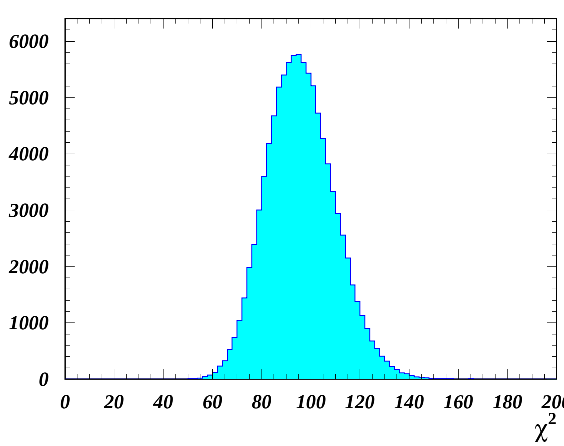

Figure 3: χ^2 distribution of 10^5 fits of the model function Eq.(9) to datasets of 100 "observation" points with Gaussian variation around the model function. Since this fit has five parameters, the mean of the distribution is $n = 100 - 5 = 95$.

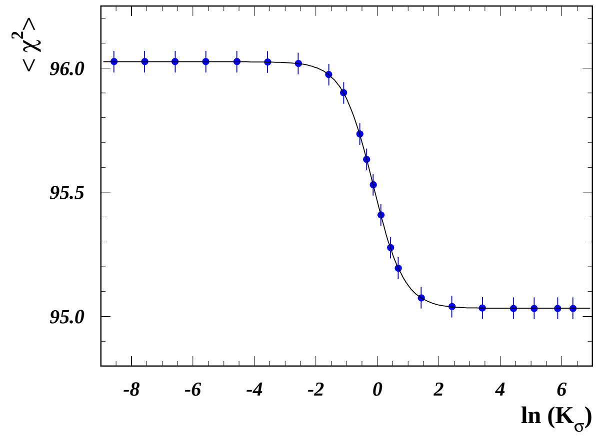

Figure 4: Evolution of the mean of the χ^2 distribution with the constraint in units of the uncertainty of the unconstrained parameter $K_\sigma = S/\sigma_0$. To clearly show the functional dependence, $\ln K_\sigma$ is plotted. Each data point is derived from a distribution like the one shown in Fig. 3. The errors represent the statistical uncertainty and are 100 % correlated. The curve is a fit of the extended $\langle\chi^2_{\text{min}}\rangle_E$ defined in Eq.(4).

4 SUMMARY

The counting of degrees of freedom is a common problem in physics. In the presence of global constraints, the number of DoF will be modified and take on a value depending on the strength of the constraint. This number need not necessarily be an integer. In this paper, a measure was proposed for the effective number of DoF in the matching of a beam line. Furthermore, a method was presented to naturally include the modification in the number of DoF in a least squares fit, based on the mean of a χ^2 distribution.

5 ACKNOWLEDGEMENTS

Sincere thanks to all who have contributed to the presented work in stimulating discussions and in particular to M. Martini and M. Schmelling for their interest and support.

6 REFERENCES

[1] H. Grote and F. C. Iselin. The MAD Program, User's Reference Manual. SL Note 90-13 (AP) (Rev. 4), CERN, March 1995.

[2] M. Martini M. Giovannozzi and A.-S. Müller. A Low-β Stripping Insertion in the CERN PS to SPS Transfer Line for the LHC Ion Programme. In *Proceedings of the 19th Particle Accelerator Conference*, 2001.

[3] A.-S. Müller and J. Wenninger. Observation of Non-Linear Effects in the Synchrotron Tune at LEP. In *Proceedings of the 9th European Particle Accelerator Conference*, 2002.