Gedanken zur Revision der Gauß'schen Fehlerrechnung

On the Revision of the Gaussian Error Calculus

Michael Grabe, Braunschweig

Manuskripteingang: 5. Juni 1999; Überarbeitung eingegangen: 8. Februar 2000; zur Veröffentlichung angenommen: 4. April 2000

Vor etwa zwei Jahrzehnten stellten die Metrologen fest, dass die Formalismen der klassischen Gauß'schen Fehlerrechnung ausgedient hatten. Ursache dafür war der zwischenzeitlich nicht mehr vernachlässigbare Einfluss sogenannter unbekannter systematischer Messfehler. – Wie bekannt, verarbeiten die Gauß'schen Formalismen allein zufällige Messfehler. Im Sinne der Forderung, Messresultate national und international vergleichbar zu halten, erhob sich seitens der Nationalen Metrologischen Staatsinstitute die Notwendigkeit, ein neues, weltweit einheitlich zu praktizierendes Verfahren zum Schätzen von Messunsicherheiten festzulegen. Welche Struktur sollte die revidierte Fassung der Gauß'schen Fehlerrechnung haben? Die Hauptfrage zielte und zielt noch immer auf das Problem, wie jene unbekannten systematischen Messfehler zu formalisieren seien. Während sich die Nationalen Metrologischen Staatsinstitute dafür entschieden, die klassische Form der Gauß'schen Fehlerrechnung weitgehend beizubehalten, soll im Gegensatz dazu an dieser Stelle ein rigoroser Neuaufbau der Gauß'schen Formalismen vorgestellt und diskutiert werden. Vorausgesetzt werden dabei stationäre Messprozesse und normalverteilte zufällige Fehler.

Two decades ago metrologists realized that the classical formalisms of Gaussian error calculus were obsolete. This is due to so-called unknown systematic errors whose influence turned out not to be negligible any more. As is well known Gaussian error formalisms exclusively process random errors. As measurements should be kept nationally and internationally comparable all the National Metrological Institutions had to standardize new procedures worldwide to be applied uniformly for the estimation of measurement uncertainties.
How should the revised Gaussian error calculus be structured? The main question referred and still refers to the problem how those unknown systematic errors should be formalized. While the National Metrological Institutes decided to essentially preserve the classical kind of Gaussian error calculus, in contrast to that, here, a rigorous revision of the Gaussian formalisms shall be presented and discussed. Thereby, stationary measurement processes and normally distributed random errors are presupposed.

Schlagwörter: Gauß'sche Fehlerrechnung, unbekannte systematische Fehler, Messunsicherheiten, GUM

1 Einleitung

Die von *K. F. Gauß* entwickelten Methoden zum Auswerten von Messungen kennen und verarbeiten allein zufällige Messfehler. Zwar erkannte Gauß die heute so relevanten *unbekannten systematischen Messfehler* – dessen ungeachtet nahm er sie in seine Formalismen nicht auf.

Im Laufe der letzten Jahrzehnte indessen ist – auf Grund verbesserter Messtechniken – der Einfluss zufälliger Messfehler soweit gesunken, dass die Tragweite *unbekannter systematischer Messfehler* nicht länger unberücksichtigt bleiben konnte. Im Falle eines zeitlich stabil arbeitenden, stationären Messprozesses sind Messfehler letzterer Art zeitlich konstante, nach Betrag und Vorzeichen unbekannte

Störgrößen. Obwohl dieser Umstand den Metrologischen Staatsinstituten – in deren Aufgabenbereich die Realisierung der physikalischen Einheiten und die metrologische Überwachung aller Transfers fallen, die Waren gegen Geld umsetzen – durchaus bekannt war, gelang es lange Zeit nicht, einheitliche Formalismen zur Verarbeitung jener, sagen wir „Messfehler zweiter Art", zu vereinbaren. Unabweisbar blieb zunächst lediglich die Einsicht, dass die klassischen Formalismen zum Auswerten von Messdaten und zum Schätzen von Messunsicherheiten ausgedient hatten.

In der Standardliteratur finden unbekannte systematische Messfehler nach wie vor keine Berücksichtigung, und was überhaupt an Behandlungsvorschlägen existierte, stand verstreut in der Spezial-Literatur und wurde dort uneinheitlich abgehandelt – eben im Sinne wissenschaftlicher Diskussionsbeiträge. Sind aber schon die Formalismen zum Schätzen von Messunsicherheiten uneinheitlich, so gestalten sich die ohnehin schwierigen nationalen und internationalen metrologischen Vergleiche noch schwieriger.

2 Historische Entwicklung

In Konsequenz der misslichen Situation, internationale metrologische Vergleiche kaum mehr befriedigend durchführen zu können, legten alle dem Bureau International des Poids et Mesures *(BIPM)* angegliederten Nationalen Metrologischen Staatsinstitute die von ihnen benutzten Verfahren schriftlich nieder. Man verglich sie und filterte aus allen schließlich den sogenannten ***Guide to the Expression of Uncertainty in Measurement*** *(GUM)* heraus, [7].

Im Grunde genommen schreibt der *GUM* die klassische Gauß'sche Fehlerrechnung fort: Er „rettet" sie, indem er unbekannten systematischen Messfehlern einen quasi-zufälligen Status verleiht, [9; 13; 14]. Formal verfügt der *GUM* damit, so wie die klassische Gauß'sche Fehlerrechnung auch, über erwartungstreue Schätzer. Indessen ist dieser Vorgehensweise schon immer vehement widersprochen worden, [5], in letzterer Zeit auch in [12].

Mit dem Erscheinen des *GUM* im Jahre 1993 ist davon auszugehen, dass Messunsicherheiten in der Tat weltweit einheitlich geschätzt werden, was zweifelsohne von Vorteil ist: Seine Formalismen sind überall anerkannt und rechtlich gesehen auch verbindlich. Damit galt das Problem Messunsicherheit als bewältigt.

Von jenen Strömungen des Vereinheitlichens unabhängig, jedoch zeitlich parallel, hat der Verfasser einen anderen Weg zur Revision der Gauß'schen Formalismen beschritten, [3; 4]. Ziel dieser Bemühungen war es, den Gauß'schen Formalismus so zu fassen, dass er in der Lage sei, *nicht-erwartungstreue Schätzer* zu verarbeiten. Diese Idee befand sich, als die Frage zur Debatte stand, welche Struktur die international einheitlich festzulegenden Formalismen zum Schätzen von Messunsicherheiten haben sollten, in ihren Anfangsgründen. Überdies war sie – gemessen an der Vorlaufzeit des GUM – offenbar auch zu neu, um sich gegen die Integration unbekannter systematischer Fehler im Sinne des Beibehaltens der klassischen Gauß'schen Formalismen durchsetzen zu können.

Im Folgenden soll erläutert werden, worin sich die beiden Ansätze unterscheiden. Dabei stützt sich der Verfasser auf die innerhalb der Angewandten Statistik üblichen Nomenklaturen. Insbesondere zieht er den Begriff *Fehler* dem im *GUM* vorgeschlagenen Begriff *Abweichung* vor, da es seiner Ansicht nach mehr eine Frage der Zielsetzung als eine Frage des physikalischen Vorgangs ist, ob eine Größe als Fehler oder ob sie als Abweichung in Erscheinung tritt: Das Ziel des Experimentators ist *der wahre Wert* der zu messenden Größe. Jeder danebenliegende Wert ist daran gemessen notwendigerweise falsch – dass dem ein physikalischer Vorgang zugrunde liegt, ist selbstverständlich und signalisiert aus der Sicht des Verfassers zumindest nicht die Notwendigkeit einer Umbenennung.

3 Messunsicherheiten erwartungstreuer Schätzer

Arbeitet der Messprozess stationär, was sich jeder Experimentator wünscht, so sind unbekannte systematische Fehler ihrem Wesen nach vor allem erst einmal zeitlich konstante Größen. Da sie nach Betrag und Vorzeichen unbekannt sind, sieht sich der *GUM* berechtigt, sie mittels wahrscheinlichkeitstheoretischer Methoden zu formalisieren. Letztlich mit dem Resultat, dass zeitlich konstanten Größen Varianzen zugeordnet werden können. Gehen wir davon aus, diese Vorgehensweise sei mathematisch begründbar. Dann aber darf dessen ungeachtet gefragt werden, inwieweit sich der zugehörige mathematische Apparat an den tatsächlichen Bedürfnissen der Metrologie orientiert – insbesondere, ob der Gewinn den Aufwand rechtfertigt. Denn die wahrscheinlichkeitstheoretische Formalisierung unbekannter systematischer Fehler provoziert aus Sicht des Verfassers Probleme.

Zunächst einmal könnte sich der Experimentator auf den Standpunkt stellen, seine Messapparatur sei ein Unikat, und so habe er nicht die Möglichkeit, auf ein Ensemble von Messapparaturen zurückzugreifen. Folglich dürfte ihm die Varianz eines unbekannten systematischen Fehlers fremd bleiben. Näher läge es schon, stattdessen mit dem Absolutbetrag jenes Fehlers zu arbeiten.

Betrachten wir ein arithmetisches Mittel aus n Wiederholungsmessungen: Physikalisch gesehen streuen die Wiederholungsmessungen um den Erwartungswert eben dieses Mittels. Ist dieser Erwartungswert aufgrund der Einflussnahme eines unbekannten systematischen Messfehlers zeitkonstant gegen den wahren Wert der Messgröße verschoben, so sollte dieses Geschehen nicht durch eine per Postulat definierte Verteilungsdichte aufgehoben werden, und zwar auch dann nicht, wenn jene Verschiebung *unbekannt* ist. Dessen ungeachtet sind die Erwartungswerte des *GUM* – eben vermöge jener postulierten Verteilungsdichte – erwartungstreu. Ein solcher Widerspruch zur physikalischen Realität des Experimentes dürfte schwerlich tolerierbar sein. In Konsequenz dieser Beurteilung des Geschehens kann der *GUM u.U.* zu niedrige Messunsicherheiten festlegen – sofern Messunsicherheiten so interpretiert werden, dass sie in der Lage sind, die wahren Werte der zu messenden physikalischen Größen, wenn nicht mit mathematischer Sicherheit so doch wenigstens quasi-sicher, zu lokalisieren. – Dieses letztere Verständnis von Messunsicherheit ist nach Ansicht des Verfassers physikalisch gesehen unverzichtbar. Im *GUM* findet es sich leider nicht. Damit steht der *GUM* – jedenfalls nach Meinung des Verfassers – an der Forderung nach physikalischer Objektivität. Wir werden diesen Gesichtspunkt aber noch genauer untersuchen.

Wenn demgegenüber hervorgehoben wird, die gegebenenfalls zu niedrig ausfallenden Messunsicherheiten des *GUM* machten *metrologische Diskrepanzen* sichtbar, und das sei eben Gewinn und Vorteil „knapper" Messunsicherheiten, so muss dem entgegengehalten werden, dass Diskrepanzen, die aufgrund zu niedriger Messunsicherheiten künstlich hervorgerufen worden wären, keine physikalischen Inhalte reflektierten. Von Diskrepanzen könnte dann gesprochen werden, wenn die wahren Werte aller in Frage stehender physikalischer Größen mittels der hier für unverzichtbar erachteten Auffassung von Messunsicherheit lokalisiert worden wären – andernfalls nämlich würden die Relationen „schwimmen".

Ob Messunsicherheiten korrekt geschätzt worden sind, lässt sich zwar nicht am Experiment selbst zeigen, denn man kennt die wahren Werte der Messgrößen nicht, wohl aber sind Rechnersimulationen in der Lage, diesen Nachweis zu führen, und sie schließen, wie bereits betont, den Effekt ein, dass die Messunsicherheiten des *GUM* möglicherweise der Forderung nach physikalischer Objektivität nicht genügen.

Bemerkenswerterweise formuliert der *GUM* aber auch keine Aussage im Sinne elementarer physikalischer Objektivität, d.h. hinsichtlich der so entscheidenden Frage, ob der Komplex aus Schätzer und Messunsicherheit den wahren Wert der Messgröße mit Sicherheit lokalisiert. Konkret sieht der *GUM* zwei Möglichkeiten vor: Entweder man arbeite mit der seinerseits definierten *1σ-Standard-Messunsicherheit* – die nach hier vertretener Auffassung in der Regel zu niedrig ausfallen dürfte – oder man erweitere sie „im Bedarfsfalle" durch Multiplikation mit einem sogenannten k_p-Faktor.

Da Rechnersimulationen zeigen, dass die *1σ-Standard-Messunsicherheit des GUM* den wahren Wert der Messgröße nicht zuverlässig überdeckt, scheidet sie aus Sicht des Verfassers für die Zwecke der Metrologie aller Genauigkeitsstufen aus, angefangen von den Notwendigkeiten des Kalibrierdienstes bis hin zum Festlegen der Fundamentalkonstanten der Physik. Die *Erweiterte Messunsicherheit des GUM* vergrößert die geschätzte Messunsicherheit mittels eines Hilfsfaktors, eben jenes k_p-Faktors. Jedoch teilt der *GUM* über dessen Berechenbarkeit nichts mit, sondern schlägt vor, in der Regel $k_p = 2$ zu setzen. Dieser Wert dürfte praxisgerecht sein und in dieser Eigenschaft sinnvolle Messunsicherheiten erzeugen. – Indessen sieht der Verfasser keine Empireme, die diese Wahl von k_p stützen könnten.

Wie sollen sich nach allem grundsätzliche Fragen der Art „Haben Neutrinos Masse?" entscheiden lassen, legte man derartige Formalismen zugrunde? Selbst wenn man einräumt, Antworten auf physikalische Fragen enthielten grundsätzlich immer eine Restunsicherheit, so darf doch kritisch hinterfragt werden, wie groß – quantitativ gesehen – eine derartige Restunsicherheit denn sein dürfe. Ist eine bündige Antwort von Interesse, so könnte der Faktor $k_p = 2$ helfen, um den Komplikationen zu niedrig geschätzter Messunsicherheiten aus dem Wege zu gehen. Aber dennoch stellte sich die Frage, warum gerade 2 und nicht vielleicht auch 2,5?

Die Fundamentalkonstanten der Physik legen (weitgehend) die physikalischen Einheiten fest. Daher möchten wir ihre Unsicherheiten so zuverlässig wie möglich kennen. Es mag durchaus sein, dass jene Unsicherheiten auch jetzt vielleicht schon sicher sind. Dennoch wäre es wünschenswert, das *ab initio* nachvollziehen zu können. Da Fundamentalkonstanten traditionell mit den *1σ-Standard-Unsicherheiten* des *GUM* versehen werden, wären Verschiebungen innerhalb des Systems der physikalischen Konstanten und damit der gesetzlichen Einheiten zumindest vorstellbar. Demgegenüber verlangen die Fundamentalkonstanten der Physik kategorisch *die kleinstmöglichen Unsicherheitsintervalle, die größtmögliche Sicherheiten* realisieren. Dieser Forderung genügt der *GUM* eigentlich nicht.

So gesehen stehen dem Guide neben seinem bemerkenswerten Erfolg, einheitliche Grundlagen zum Auswerten von Messungen geschaffen zu haben, eine Reihe von Nachteilen gegenüber: An erster Stelle steht die Objektivität der Metrologie, und damit die Frage der Rückführung von Messwerten auf gesetzliche Einheiten – einschließlich der Rückführung der gesetzlichen Einheiten auf sich selbst – innerhalb ihrer jeweils *eigenen* Hierarchien.

Der *GUM* ist als Kompilation verschiedener älterer Auswerteverfahren nicht aus einem Guss – weder in seiner Darstellung noch in der ihm zugrunde liegenden Konzeption. In dieser Eigenschaft teilt er das Schicksal anderer, aus unterschiedlichen Denkansätzen und Quellen zusammengestellter Extrakte: Sie sind schwierig zu lesen und schwierig zu handhaben.

4 Messunsicherheiten nichterwartungstreuer Schätzer

Bemerkenswerterweise sind die Grundideen der in diesem Beitrag noch einmal vorgestellten, mit nichterwartungstreuen Schätzern arbeitenden Fehlerrechnung älter als der *GUM*. Wie der Verfasser schon 1978 zeigen konnte, [3], lassen sich – sofern die Messprozesse stationär sind – unbekannte systematische Messfehler als zeitkonstante Größen formalisieren und in dieser Eigenschaft konsequent von den zufälligen Messfehlern trennen. Dementsprechend ließ sich auch die Fortpflanzung der beiden so unterschiedlichen Messfehlertypen trennen. Allerdings um den Preis neuer Formalismen der Fehlerrechnung.

Dieser damals noch unvollständige Ansatz erwies sich als Urversion einer rigorosen Neufassung der Gauß'schen Fehlerrechnung. Nach Abschluss der Revisionsarbeiten sieht sich der Verfasser ermutigt, das neue Konzept mit dem *GUM* zu vergleichen. – Einzelne Abschnitte der neuen Fehlerrechnung sind zwar publiziert worden, so in [5; 8; 11], jedoch erwies sich die Gesamtdarstellung des Konzeptes, [15], letztlich als zu umfangreich, um der kompakten Form von Zeitschriftenartikeln genügen zu können.

Die neue, mit nichterwartungstreuen Schätzern operierende Fehlerrechnung belässt unbekannten systematischen Messfehlern ihre Basiseigenschaft, physikalische Konstante zu sein. Dieses Bild trägt einerseits den Fakten der physikalischen Messung Rechnung – immer stationäre Messprozesse vorausgesetzt – andererseits lässt sich mit ihm die Forderung nach Definition sicherer, nachvollziehbarer Messunsicherheiten realisieren. Die folgenden Betrachtungen setzen normalverteilte zufällige Messfehler voraus. Da zufällige und systematische Messfehler voneinander getrennt formalisiert werden, kann das neue Konzept dann auf die gängigen Techniken der Angewandten Statistik zurückgreifen, was sich als außerordentlich vorteilhaft erweist: Die zufälligen Fehler definieren konventionelle Vertrauensbereiche, zu denen die mittels *modifizierter* Worst-Case-Abschätzungen auf der Basis der Dreiecksungleichung verknüpften und fortgepflanzten unbekannten systematischen Fehler linear hinzuaddiert werden.

Die Fehlerrechnung der Physiker und Ingenieure kennt im Rahmen der Fortpflanzung zufälliger Messfehler keine Vertrauensbereiche. Der Grund hierfür ist leicht zu sehen: Nehmen wir an, die Messgrößen X und Y seien zu verknüpfen. Stammen X und Y aus verschiedenen Laboratorien, werden die Zahlen n_x und n_y der jeweiligen Wiederholungsmessungen in der Regel unterschiedlich ausfallen. „Um keine eventuell nützliche Information zu verlieren", ist es Brauch, jene *unterschiedlich vielen* Wiederholungsmessungen in die Auswerteprozeduren einfließen zu lassen. Wie man sich überlegt, macht eben diese Vorgehensweise die Möglichkeit zunichte, auf einfacher, sinnvoller Basis Vertrauensbereiche zu definieren. Leider ist das Arbeiten mit unterschiedlich vielen Wiederholungsmessungen fester Bestandteil der Standardliteratur. Abhilfe schafft die Forderung nach *gleich vielen* Wiederholungsmessungen für jede der zu verknüpfenden Messgrößen, [8], was sich in der Regel nur durch Streichen (!) überzähliger Messwerte erreichen lässt. Das ist erlaubt, weil das Resultat einer Messung nicht signifikant von der Zahl der Wiederholungsmessungen abhängen darf, und sinnvoll, weil das Arbeiten mit gleich vielen Wiederholungsmessungen zu den Vertrauensbereichen der Student-Dichte zurückführt, sofern die zufälligen Fehler normalverteilt sind, [1]. Zum Einwand, es wäre ja nicht entscheidbar, „welche" Messungen gegebenenfalls zu streichen wären und welche nicht, folgende Anmerkungen:

- Arbeitet der Messprozess statistisch gesehen stationär, so ist es gleichgültig, welche Messungen gestrichen werden, wenn die zu verknüpfenden Messreihen voneinander unabhängig sind. Sind sie voneinander abhängig, so werden in aller Regel ohnehin Paare von Messdaten registriert werden, so dass sich ein Streichen überzähliger Messungen erübrigt.
- So genannte Ausreißer signalisieren in der Regel experimentelle Unzulänglichkeiten der Messapparatur oder Nichtstationaritäten des physikalischen Effektes. Was die Apparatur betrifft, könnte eine Reparatur der Anlage u.U. mehr versprechen als eine Ausreißeranalyse.
- Sollte sich die hier erhobene Forderung nach gleich vielen Wiederholungsmessungen in der Praxis durchsetzen, ließe sich ein Bereich oder auch eine Mindestzahl von Wiederholungsmessungen vereinbaren.

Ein weiterer, mit der Forderung nach gleich vielen Wiederholungsmessungen in Zusammenhang stehender Aspekt betrifft die Kovarianzen zu verknüpfender Messgrößen. Wie die Verteilungsdichte der empirischen Momente zweiter Ordnung, d.h. der empirischen Varianzen und Kovarianzen, zeigt, sind die empirischen Varianzen und Kovarianzen *immer voneinander abhängig,* und zwar *auch dann*, wenn die Messgrößen selbst voneinander unabhängig oder lediglich unkorreliert sind. Dieser Umstand verlangt, die empirischen Kovarianzen grundsätzlich mitzunehmen, [8], was natürlich nur möglich ist, wenn für alle beteiligten Messgrößen *gleich viele* Wiederholungsmessungen vorliegen.

Die Standardliteratur unterscheidet traditionell nicht scharf zwischen empirischen Varianzen und Kovarianzen auf der einen Seite und den zugehörigen theoretischen Größen, d.h. den Erwartungswerten, auf der anderen. Diese Vorgehensweise ist misslich deshalb, weil dem Experimentator allein erstere, die empirischen Größen, zugänglich sind. Zu ihnen gehören spezielle Verteilungsdichten, entweder die Student'sche Dichte oder, wie hier nicht im Detail gezeigt werden kann, die seitens der Physiker und Ingenieure faktisch vergessene Hotelling'sche Dichte. Die Praxis arbeitet naturgemäß mit empirischen Größen – denn Erwartungswerte sind ihr ja nicht bekannt – und bringt dessen ungeachtet jene genannten, passenden Verteilungsdichten nicht immer zum Tragen. Beispiele für das Arbeiten mit physikalisch unzugänglichen Erwartungswerten finden sich in [2; 10]. Eine untragbare Situation wird uns im Rahmen des Ausgleichs nach kleinsten Quadraten begegnen.

Kommen wir auf die oben angesprochene Modifikation der Dreiecksungleichung zurück. Sie sorgt zum einen für die Eindeutigkeit der Verknüpfung unbekannter systematischer Messfehler. D.h., erstreckt sich die Fehlerfortpflanzung über mehrere Verknüpfungsstufen hinweg – das ist der Regelfall – und bestehen verschiedene Wege dorthin, so stellt besagte Modifikation sicher, dass der Beitrag der fortgepflanzten unbekannten systematischen Fehler stets *denselben Wert* annimmt. Zum anderen verhindert jene Modifikation das sinnlose Kumulieren systematischer Fehleranteile – denn das wäre in der Tat eine missverstandene Anwendung der Dreicksungleichung.

Nach allem gestaltet sich die so vertretene Fehlerrechnung „gläsern“: Sie arbeitet wie ein Baukasten. Jede ihrer Verknüpfungsstufen ist leicht und unmittelbar einsehbar, insbesondere sind alle Unsicherheiten reversibel – d.h., sie können rückwärtsgehend bis in beliebige Tiefe aufgeschlüsselt werden. Die Messunsicherheiten genügen der Prämisse, *die kleinsten die nichterwartungstreuen Schätzer umgebenden Intervalle zu sein, die die wahren Werte der in Frage stehenden Messgrößen mit (vernunftgemäßer) Sicherheit überdecken*. Mit (vernunftgemäßer) Sicherheit, weil Vertrauensbereiche der Vertrauenshöhe, sagen wir 95%, seitens des Experimentes nicht ausgeschöpft werden (das Experiment realisiert die Ausläufer der Normalverteilung nicht) und weil hinsichtlich des Abschätzens des Einflusses unbekannter systematischer Fehler dem ungünstigsten Fall Rechnung getragen wird.

Diese Aussage lässt sich mit Hilfe von Rechner-Simulationen stützen.

Der hier lediglich umrissene Formalismus löst oder erklärt überdies faktisch alle bisher nicht oder formal nur schwer lösbaren Probleme der Fehlerrechnung. Nehmen wir den Ausgleich nach kleinsten Quadraten.

Das Gauß-Markoff'sche Theorem setzt erwartungstreue Schätzer voraus. Andererseits ist jeder unbekannte systematische Fehler, wiewohl unbekannt, nichts anderes als ein Bias oder eine Vorlast. In Konsequenz von Vorlasten verliert das Gauß-Markoff'sche Theorem seine Gültigkeit: Die *Praxis* zeigt, dass sich die minimierte Summe der Residuenquadrate kaum je als Schätzer der Zahl der Freiheitsgrade des auszugleichenden Systems interpretieren lässt, [6]. Diese Feststellung belegt so gut wie jeder Ausgleich von Messdaten, die Träger unbekannter systematischer Fehler sind. Davon unabhängig gelingt es durchaus, sinnvolle Messunsicherheiten für die Schätzer des Ausgleichs festzulegen. Natürlich existiert dabei keine Gewichtsmatrix im Sinne des Gauß-Markoff'schen Theorems. Wie gewichtet werden kann, ist in [11] gezeigt worden.

Kommen wir zu den Korrelationen der Komponenten des Lösungsvektors: Da jede Komponente denselben Satz von Eingangsdaten einschließt, ist jede Komponente mit jeder anderen korreliert. Mittels der hier zur Diskussion gestellten Formalismen lassen sich jene Abhängigkeiten rezeptartig formalisieren und graphisch in Gestalt zwei- oder dreidimensionaler Körper darstellen, gegebenenfalls auch als dreidimensionale Schnitte von Körpern aus Räumen höherer Ordnung, [8; 11]. Vergleichbares kann der *GUM* nach Verständnis des Verfassers nicht – übrigens auch nicht die Fehlerrechnung der *„Vor-GUM-Zeit“*.

Die mathematische Statistik beschreibt die Kopplungen zwischen den Schätzern des Ausgleichs mittels des χ^2-verteilten Exponenten der mehrdimensionalen normalen Verteilungsdichte. Das ist richtig, aber eben nur in der Statistik. In der Physik ist dasselbe unzulässig – dessen ungeachtet wird es getan, [2]. Derartiges ist in der Physik unzulässig, weil der Experimentator die theoretischen Varianzen und die theoretischen Kovarianzen des χ^2-verteilten Exponenten nicht kennt. Was er kennt, sind *empirische* Größen, d.h. empirische Varianzen und empirische Kovarianzen, und für diese existiert jene wohldefinierte, in der Metrologie bis heute unbenutzt bleibende Hotelling'sche Dichte. Letztere kommt mit *empirischen Schätzern* aus und beseitigt so das Problem physikalisch nicht korrekter Konfidenz-Ellipsoide – normalverteilte zufällige Fehler vorausgesetzt.

Denken wir des Weiteren an die Varianz-Analyse: Ihre gelegentlich scheinbare Widersprüchlichkeit ist auf die den Messdaten überlagerten zeitkonstanten, unbekannten systematischen Fehler zurückzuführen. Die Varianz-Analyse ist ein Produkt der reinen Statistik. Ist, wie in der Experimentalphysik der Fall, *a priori* bekannt, dass den Messdaten unbekannte Konstanten überlagert sind, versagt die Varianz-Analyse ohne jeden weiteren Kommentar. Infolgedessen bleibt kein anderer Weg, als sich dazu zu entschließen, die Varianz-Analyse aus dem Bereich der Physik ein für allemal zu verbannen: Sie ist dort nicht anwendbar.

5 Schlussfolgerung

Die Entscheidung der Nationalen Metrologischen Staatsinstitute, die international einheitlich zu vereinbarende Fehlerrechnung auf den *GUM* zu stützen, mag aus der Zeit des Umbruchs heraus verständlich gewesen sein. Indessen könnte der Preis dieser Entscheidung hoch sein, geht es doch um nicht weniger als um die Objektivität der Naturwissenschaft.

Der seitens des Autors entwickelte, hier nur umrissene Weg, [15], hat ebenfalls seinen Preis: Der Experimentator müsste sich mit neuen Formalismen vertraut machen. Davon unabhängig gewähren letztere robuste, sichere Messunsicherheiten, deren Aussagekraft selbst in Fällen grundsätzlicher Art kaum in Zweifel zu ziehen wäre. Die neuen Messunsicherheiten sind vergleichbar mit den *2σ-Standardunsicherheiten* des *GUM*, gegebenenfalls sind sie auch größer, keineswegs aber sind sie physikalisch unangemessen groß. Die unkomplizierte, baukastenähnliche Struktur der Formalismen ließe sich auf *allen Stufen* der metrologischen Hierarchien einheitlich anwenden.

Angesichts der Leistungsfähigkeit dieser Formalismen sollten die metrologischen Staatsinstitute die 1993 getroffene Entscheidung vielleicht doch noch einmal überdenken.

Danksagung

Der Autor dankt Herrn Dr. W. Wöger für zahlreiche kritische Anmerkungen.

Literatur

[1] Graybill, F. A.: *An Introduction to Linear Statistical Models*, McGraw-Hill, New York 1961.

[2] Eadie, W. T. et al.: *Statistical Methods in Experimental Physics*, North-Holland Publishing Company, 1971.

[3] Grabe, M.: in *Seminar über das Schätzen der Messunsicherheit*, PTB, 20. und 21. Februar 1978.

[4] Grabe, M.: *On the Assignment of Uncertainties within the Method of Least Squares*, Poster Paper, Second International Conference on Precision Measurement and Fundamental Constants, Washington, D.C. 8.–12. June 1981.

[5] Grabe, M.: *Principles of „Metrological Statistics“*, metrologia, 23 (1986/87) 213–219.

[6] CODATA Task Group: *The 1986 Adjustment of the Fundamental Physical Constants,* CODATA Bulletin, 63(1986), Pergamon Press.

[7] *GUM*, *Guide to the Expression of Uncertainty in Measurement*, 1993, 1, Rue Varambe, Case Postale 56, CH-1211 Geneva 20, Switzerland. Unterstützt durch folgende Organisationen: International Bureau of Weights and Measures (BIPM), International Organisation of Legal Metrology (OIML), International Organisation for Standardisation (ISO), International Electrotechnical Commission (IEC), International Union of Pure and Applied Chemistry (IUPAC), International Union of Pure and Applied Physics (IUPAP), International Federation of Clinical Chemistry (IFCC).

[8] Grabe, M.: *On the estimation of one- and multi-dimensional uncertainties*, National Conference of Standard Laboratories, Albuquerque, N.M. (USA), Proceedings 1993, 569–576.

[9] Weise, K. and Wöger, W.: *A Bayesian theory of measurement uncertainty*, Meas. Sci. Technol. 4 (1993) 1–11.

[10] Bich, W.: Simple formula for the propagation of variances and covariances, metrologia 33 (1966) 181–183.

[11] Grabe, M.: An Alternative Algorithm for Adjusting the Fundamental Physical Constants, Physics Letters A 213 (1996) 125–137.

[12] Bich, W.: *The ISO Guide to the Expression of Uncertainty in Measurement: A Bridge Between Statistics and Metrology*, in Advanced Mathematical Tools in Metrology III, Eds. P. Ciarlini, M. G. Cox, F. Pavese, D. Richter., World Scientific 1997.

[13] Kessel, W.: European and international standards for statements of uncertainty, Eng. Sci. Educat. J. 7 (1998) 201–207.

[14] Weise, K. und W. Wöger: *Messunsicherheit und Messdatenauswertung*, Wiley-VCH, 1999.

[15] Grabe, M.: Schätzen von Messunsicherheiten in Wissenschaft und Technik, Manuskript, 300 S.

Hinweis des Herausgebers: Die vorliegende Arbeit bringt die persönliche Ansicht des Autors zum Ausdruck. Technisches Messen wird dem *Guide to the Expression of Uncertainty in Measurement (GUM)* und dessen Umsetzung in die Praxis das Heft 11.2000 widmen.

Dr. Michael Grabe, Am Hasselteich 5, D-38104 Braunschweig, Tel. / Fax: 0531-371642, E-Mail: michael.grabe@ptb.de