

FP2 Teilchenphysik

Das Energiespektrum des Krebsnebels, gemessen mit FACT

Astroteilchenphysik, TU Dortmund

15. April 2021

Inhaltsverzeichnis

1 Abstract	3
2 Aufgaben: Anwendung auf FACT	3
2.1 Download der Daten	3
2.2 Daten-Selektion	3
2.3 Theta-Quadrat-Plot und Detektions-Signifikanz	3
2.4 Energie-Migration	4
2.5 Entfaltung	4
2.6 Akzeptanz-Korrektur und Berechnung des Flusses	4
2.7 Vergleich mit MAGIC und HEGRA	4
3 Bogengebundene Gammaastronomie	5
4 FACT	5
4.1 Beobachtungsmodus	7
5 Datensatz	7
6 Entfaltung	10
6.1 Naive SVD-Entfaltung	10
6.2 Poisson-Likelihood-Entfaltung	11
6.3 Iterative Bayes Unfolding	11
6.4 Regularisierung	11
7 Hinweise	12
7.1 Programmier-Hinweise	12
Literatur	12

1 Abstract

In diesem Versuch soll das Energiespektrum der vom Krebsnebel (M1) ausgesandten Gamma-Strahlung im Energiebereich zwischen 500 GeV und 15 TeV bestimmt werden.

Hierfür werden rekonstruierte Ereignisse, gemessen bzw. simuliert für das First G-APD Cherenkov Telescope verwendet.

Es sollen verschiedene Entfaltungsalgorithmen verwendet werden, um von den rekonstruierten Gammaereignissen das Energiespektrum zu berechnen.

2 Aufgaben: Anwendung auf FACT

Dies ist eine Kurzfassung der durchzuführenden Aufgaben, Hintergrundinformationen zu den Daten und den zu verwendenden Methoden gibt es in den weiteren Kapiteln.

2.1 Download der Daten

Für die Bestimmung des Energiespektrums des Krebsnebels auf dem öffentlichen FACT Datensatz werden drei Datensätze benötigt:

open_crab_sample_dl3.hdf5 Rekonstruierte Messdaten aus 17,7 h Observationszeit des Krebsnebels.

gamma_test_dl3.hdf5 Rekonstruierte, simulierte Gamma-Ereignisse

gamma_corsika_headers.hdf5 Informationen über alle simulierten Gamma-Luftschauer.

Download von <https://factdata.app.tu-dortmund.de/fp2/>

2.2 Daten-Selektion

Wählen Sie aus den rekonstruierten Ereignissen (gemessen sowie simuliert), nur die Ereignisse aus, die vom Random-Forest-Classifizier mit einer Konfidenz von ≥ 0.8 als Gamma klassifiziert wurden. Die Konfidenz des Random-Forest-Classifiers ist in der Spalte `gamma_prediction` gespeichert.

2.3 Theta-Quadrat-Plot und Detektions-Signifikanz

Erzeugen sie für die gemessenen Krebsnebel-Daten den Theta-Quadrat-Plot, siehe Abbildung 5. Hier werden für die Signal-Region, sowie die 5 Off-Regionen kombiniert, das Quadrat des Abstands der rekonstruierten von der vermuteten Quellposition histogrammiert.

Wählen Sie alle Ereignisse als On bzw. Off aus, die einen maximalen Abstand zur Quellposition bzw. zu einer der Off-Positionen von $\theta^2 \leq 0,25^\circ$ haben. Der Abstand zur Quellposition ist in der Spalte `theta_deg` und der Abstand zur i -ten Off-Position in `theta_deg_off_{i}` gespeichert.

Berechnen Sie die Detektions-Signifikanz für den Krebsnebel nach Li & Ma, (3).

2.4 Energie-Migration

Bestimmen Sie aus den simulierten Gamma-Ereignissen die Detektorantwortmatrix \mathbf{A} . Im vorliegenden Fall ist dies die sogenannte Energie-Migrationsmatrix des Random Forest Regressors. Die Spalten von Interesse sind `gamma_energy_prediction` für die geschätzte Energie des Random Forest und `corsika_event_header_total_energy` für die wahre Energie des simulierten Gammas.

Verwenden Sie Bins im Bereich zwischen 500 GeV und 15 TeV, diese sollten logarithmisch-äquidistant sein. Inverse Probleme sind oft besser konditioniert, wenn mehr Bins in der gemessenen Größe als in der zu entfaltenden Größe gewählt werden. Es müssen in der Regel auch Ereignisse berücksichtigt werden, die außerhalb des Energiebereichs von Interessen liegen.

2.5 Entfaltung

Entfalten Sie die Messung des Krebsnebel-Datensatzes einmal mit der naiven SVD-Entfaltung und einer weiteren Methode ihrer Wahl.

2.6 Akzeptanz-Korrektur und Berechnung des Flusses

In der Gamma-Astronomie wird die Akzeptanz eines Detektors als die effektive Fläche eines perfekten Detektors angegeben. Das vollständige Experiment ist am Ende ein einfacher Binomialer Prozess, bei dem die Wahrscheinlichkeit $p(E)$ die Wahrscheinlichkeit ist, ein Ereignis der wahren Energie E zu detektieren. Die effektive Fläche ist $p(E) \cdot A$, wobei A die vom Experiment abgedeckte Fläche ist. Die effektive Fläche muss für jedes Bin in wahrer Energie bestimmt werden, aus den simulierten und gemessenen Ereignissen.

$$A_{\text{eff},i} = \frac{N_{\text{selektiert},i}}{N_{\text{simuliert},i}} \cdot A \quad (1)$$

Die Gammaereignisse wurden in einem Radius von 270 m simuliert.

Berechnen sie aus den entfalteten Ereigniszahlen \hat{f}_i den Fluss des Krebsnebels:

$$\Phi_i = \frac{\hat{f}_i}{A_{\text{eff},i} \cdot \Delta E_i \cdot t_{\text{obs}}} \quad (2)$$

Wobei ΔE_i die Breite des Energiebins ist, und t_{obs} die Dauer der Beobachtung. Diese kann aus der Spalte `ontime` der `runs` Tabelle berechnet werden.

Der Datensatz `gamma_test_dl3.hdf5` enthält nur eine zufällige Untermenge von 70 % der simulierten Ereignisse, dies muss bei der Bestimmung der Akzeptanz berücksichtigt werden.

2.7 Vergleich mit MAGIC und HEGRA

Vergleichen Sie ihr Ergebnis mit den Fits an das Krebsnebelspektrum, die die MAGIC [3] und HEGRA [2] Kollaborationen veröffentlicht haben. Stellen Sie hierzu ihre entfalteten Flusspunkte in der gleichen Abbildung dar, wie die Ausgleichsfunktionen, die den oben genannten Veröffentlichungen entnommen werden müssen.

3 Bogengebundene Gammaastronomie

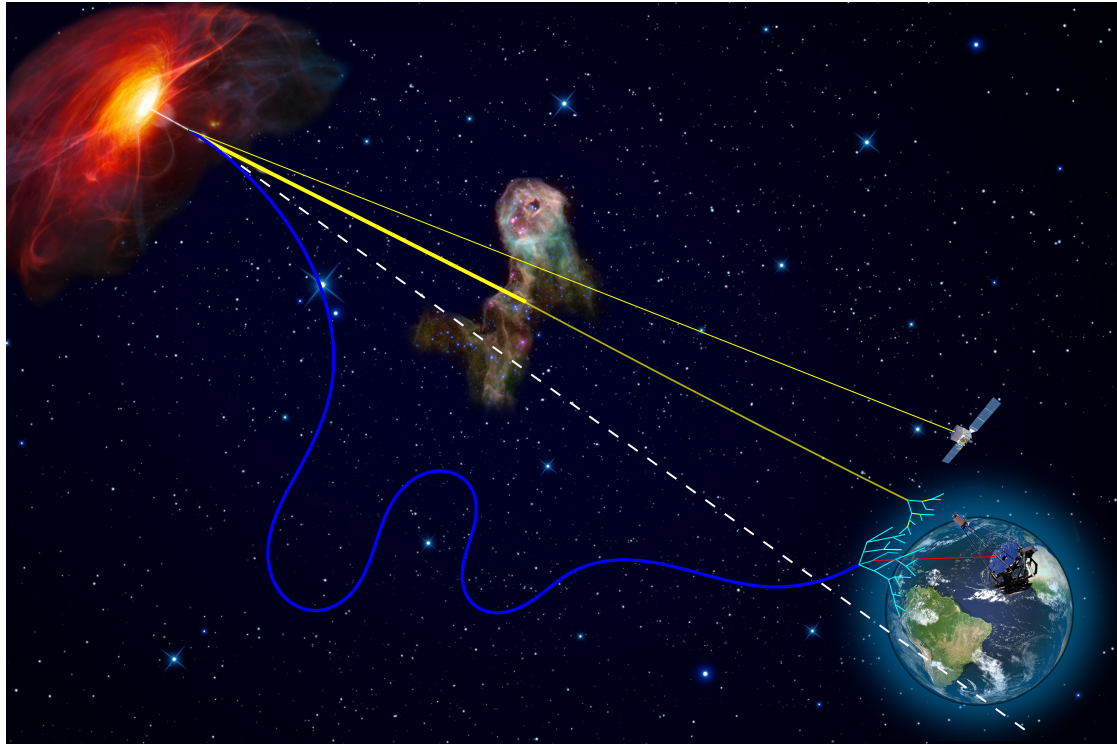


Abbildung 1: Überblick über die Botenteilchen der Astroteilchenphysik, insbesondere der Gammaastronomie. Geladene Kosmische Strahlung (blau), wird durch Magnetfelder im intergalaktischen und interstellarem Medium abgelenkt und lässt sich deshalb nicht zu ihrer Quelle zurückverfolgen. Neutrinos (weiß) sind nur extrem schwierig nachzuweisen, einzelne Punktquellen von Neutrinos wurden noch nicht entdeckt, es gibt jedoch starke Hinweise [1]. Hochenergetische Photonen können entweder direkt über Satelliten wie FERMI nachgewiesen werden oder indirekt über bogengebundene Tscherenkow-Teleskope.

4 FACT

FACT, das *First G-APD Cherenkov Telescope* [4, 5], ist ein bogengebundenes, abbildendes Tscherenkow-Teleskop, das seit 2011 im Observatorio del Roque de los Muchachos operiert. Ziel ist es eine für die Gammaastronomie neue Detektortechnologie, siliziumbasierte Halbleiterphotodetektoren, zu testen und die hellsten Gammastrahlungsquellen möglichst lückenlos zu überwachen. Neben der Standardkerze der Gammaastronomie, dem Krebsnebel, einem Supernovaüberreset in der Milchstraße, beobachtet FACT hauptsächlich extra-galaktische Quellen, sogenannte Blazare. Seit 2012 wird FACT ferngesteuert, seit 2017 operiert es robotisch.



Abbildung 2: FACT im Observatorio del Roque de los Muchachos, La Palma, Spanien. Im Hintergrund ist das MAGIC-1 Teleskop und das William-Herschel-Telescope zu sehen. [M. Nöthe, 2018]

4.1 Beobachtungsmodus

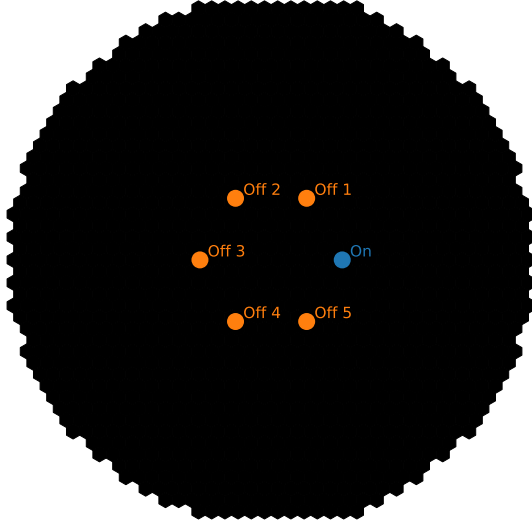


Abbildung 3: Das FACT Kamerabild mit der On-Position und den 5 Off-Positionen.

FACT beobachtet Quellen im sogenannten Wobble-Modus, um gleichzeitig mit der Quellbeobachtung auch eine Abschätzung der Untergrundrate durchzuführen. Hierzu wird mit dem Teleskop $0,6^\circ$ neben die Quellposition gezielt. Dies hat den Effekt, dass im Kamerabild nun zur Quellposition geometrisch äquivalente Positionen für die Untergrundabschätzung definiert werden können. Die Quellregion wird als *On*-Region, die Regionen zur Untergrundabschätzung werden als *Off*-Regionen bezeichnet. Dies ist in Abbildung 3 dargestellt.

Die Anzahl der rekonstruierten Ereignisse in der On-Region wird als N_{on} , die in der Off-Region als N_{off} bezeichnet. Die Signifikanz der Detektion kann mithilfe Likelihood-Ratio-Tests nach Li & Ma[11, (17)] berechnet werden:

$$S = \sqrt{2} \cdot \sqrt{N_{\text{on}} \ln \left(\frac{1 + \alpha}{\alpha} \left(\frac{N_{\text{on}}}{N_{\text{on}} + N_{\text{off}}} \right) \right) + N_{\text{off}} \ln \left((1 + \alpha) \left(\frac{N_{\text{off}}}{N_{\text{on}} + N_{\text{off}}} \right) \right)} \quad (3)$$

Hierbei ist α das Größenverhältnis der On- zur Off-Region, bei 5 Off-Regionen im Wobble-Modus gilt also $\alpha = 0,2$.

5 Datensatz

Für diesen Versuch werden bereits rekonstruierte Ereignisse verwendet. Dieser Abschnitt erklärt, welche Analyseschritte bereits durchgeführt wurden, um den Datensatz dieses Versuchs zu erzeugen.

Für jedes von bogengebundenen Tscherenkow-Teleskopen aufgenommen Schauer-Ereignis müssen drei Eigenschaften des ursprünglichen Teilchens rekonstruiert werden:

Teilchenklasse Klassifikation, üblicherweise Gamma/Hadron.

Energie Regression, meist nur für die Gamma-Kandidaten durchgeführt

Herkunftsrichtung 2d-Regression, entweder in Himmels- oder Detektorkoordinaten.

Die klassische Analyse von FACT-Daten bis zur Rekonstruktion dieser Eigenschaften besteht aus folgenden Schritten:

1. Kalibration der Spannungszeitreihen jedes Pixels
2. Ermittlung der Anzahl Photonen und deren mittlere Ankunftszeit in jedem Pixel
3. Auswahl von Pixeln, die Tscherenkow-Signal enthalten
4. Parametrisierung des Schauerbildes

Diese Schritte werden von der Software **FACT-Tools** durchgeführt. Für die hier vorliegenden Daten wurde die Version 1.1.2 genutzt [7].

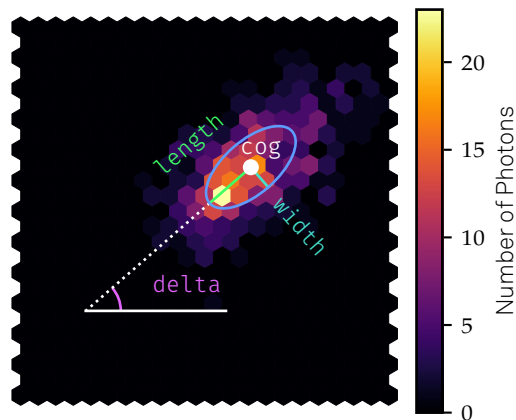


Abbildung 4: Ein Schauerbild mit Anzahl Photonen in jeden Pixel mit den wichtigsten Hillas-Parametern. Eine Hauptkomponentenanalyse auf der 2d-Lichtverteilung ergibt die beiden Standardabweichungen **length** und **width**, der Mittelpunkt wird als **cog** bezeichnet. Die Orientierung der Hauptachse wird mit dem Winkel **delta** zur x -Achse beschrieben.

Zur Bestimmung der Teilchenklasse, der Energie und Richtung sowie zur Bestimmung der Detektor-Akzeptanz sind aufwendige Simulationen nötig, um die benötigten Trainingsdaten zu erzeugen. Die Luftschauer und Tscherenkow-Produktion wurden mit der Software **CORSIKA**[10] durchgeführt. Eingangswerte für **CORSIKA** sind die Eigenschaften des Primärteilchens. **CORSIKA** propagiert nun die Teilchen durch die Atmosphäre und speichert das den Detektor erreichende Tscherenkow-Licht.

Anschließend wird mit der Software **CERES** das FACT-Teleskop simuliert. Ausgehend von den Tscherenkow-Photonen am Boden, werden die Spiegel, die Kamera-Elektronik, der Trigger und die Datennahme simuliert und zuletzt Dateien erzeugt, die das gleiche Format haben, wie die vom echten Teleskop geschriebenen Daten. Da bei diesen Datensätzen jedoch die ursprünglichen Informationen über das Primärteilchen bekannt sind, können sie für das Training von Modellen des maschinellen Lernens genutzt werden.

Es wurden drei verschiedene Datensätze simuliert:

- Gammas, aus einer punktförmigen Quelle, die im Wobble-Modus beobachtet wird
- Gammas, die diffus aus zufälligen Orten im Blickfeld des Teleskops kommen

- Protonen, die diffus aus zufälligen Orten im Blickfeld des Teleskops kommen

Als Datensatz zur Bestimmung der Detektor-Akzeptanz wurden 70 % der simulierten Quellgammas abgesondert. Dies ist der in diesem Versuch benutzte Datensatz `gamma_test_d13.hdf5`.

Der Energieschätzer, in diesem Fall ein Random Forest Regressor, nutzt die Quell-Gammas als Trainingsdaten. Die Teilchenart wird mithilfe eines Random Forest Klassifikator, der darauf trainiert wurde, die diffusen Gammas von den Protonen zu unterscheiden, klassifiziert. Für die Richtungsrekonstruktion wird die sogenannte `disp`-Methode genutzt, hier wird die 2d-Regression für den Herkunftsort der Teilchen durch die Annahme vereinfacht, die Quelle läge auf der Schauerhauptachse. Durch diese Annahme reduziert sich das Problem auf eine 1d-Regression, den Abstand der Quelle auf der Hauptachse vom Schwerpunkt des Schauers und eine binäre Klassifikation, in welcher der beiden möglichen Richtungen auf Hauptachse die Quelle liegt. Für den hier vorliegenden Datensatz wurden ein Random Forest Regressor und ein Random Forest Klassifikator auf den diffusen Gammas trainiert.

Das Ergebnis dieser Analyse[12], verfügbar unter[9], sind zwei Datensätze:

- `gamma_test_d13.hdf5` Testdatensatz bestehend aus 70 % der ursprünglich simulierten Quell-Gammas.
- `open_crab_sample_d13.hdf5` Messdaten aus 17,7 h Observationszeit des Krebsnebels.

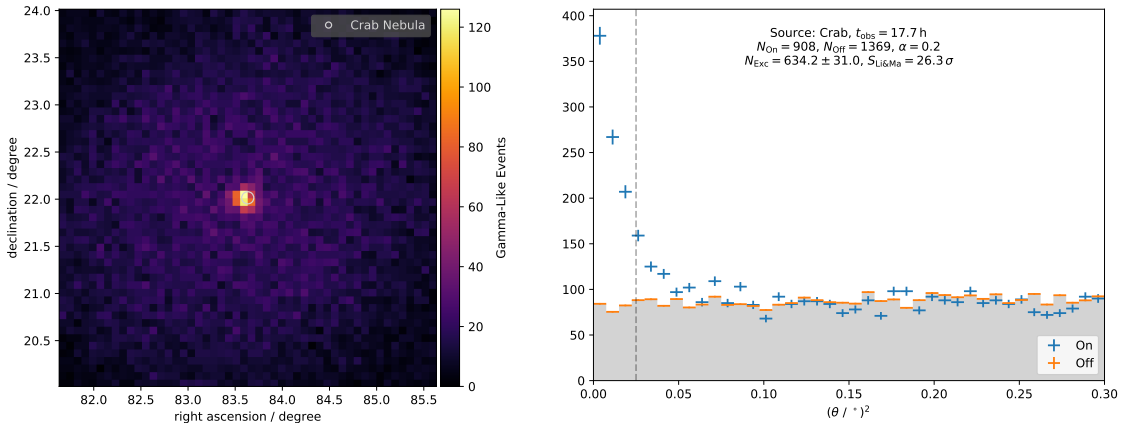


Abbildung 5: Himmelskarte der Gammakandidaten (links) und Theta-Quadrat-Plot (rechts). Die Himmelskarte zeigt alle Datenereignisse, die mit einer Konfidenz von $\geq 0,8$ vom Random Forest Klassifikator als Gamma klassifiziert wurden. Der Theta-Quadrat-Plot zeigt den Abstand von der rekonstruierten Quellposition zur angenommenen Quellposition, einmal für den Krebsnebel (On) und für 5 Untergrundregionen (Off). In dieser Messung wurde der Krebsnebel mit einer statistischen Signifikanz von $26,3 \sigma$ gemessen.

Ziel dieses Versuches ist es nun Anhand der rekonstruierten Ereignisse das Energiespektrum des Krebsnebels zu schätzen. Hierfür muss aus der Verteilung der gemessenen Energien und den Detektoreigenschaften auf das ursprüngliche emittierte Spektrum geschlossen werden.

6 Entfaltung

In physikalischen Messprozessen stehen die Größen von Interesse so gut wie nie für eine direkte Messung zur Verfügung. Gerade in der (Astro-)Teilchenphysik sind indirekte Messmethoden, bei den meistens von Ladungsdepositionen in Detektoren zu bestimmten Zeiten auf die ursprünglichen Eigenschaften der Teilchen zurückgeschlossen werden muss, die Regel.

Von der Verteilung der den Messungen zugänglichen Größen auf die zugrundeliegenden Verteilung der physikalischen Größen zu schließen wird als *inverses Problem* bezeichnet. Dies lässt sich als Fredholmsche-Integralgleichung erster Art darstellen:

$$g(y) = \int A(y, x) f(x) dx + b(y) \quad (4)$$

Hierbei bezeichnet $f(x)$ die Wahrscheinlichkeitsdichte von Interesse, abhängig von der physikalischen Größe x , $g(y)$ ist die Wahrscheinlichkeitsdichte der gemessenen Größen y , $b(y)$ ist zusätzlicher Untergrund, und $A(y, x)$ ist der Faltungskern, der den Detektor darstellt und die physikalische Größe x in gemessene Größen y überführt. Der Faltungskern ist die bedingte Wahrscheinlichkeit $P(Y|X = x)$.

(4) muss mit geeigneten Verfahren invertiert werden, um in einem physikalischen Experiment von den gemessenen Größen auf die physikalischen Größen von Interesse zurückzuschließen.

Als erster Schritt, werden die kontinuierlichen Wahrscheinlichkeitsdichten diskretisiert, sodass sie Histogramme der jeweiligen Größen entsprechen. Hierdurch vereinfacht sich (4) zu

$$\mathbf{g} = \mathbf{A}\mathbf{f} + \mathbf{b} \quad (5)$$

mit den M -Vektoren \mathbf{g}, \mathbf{b} , dem N Vektor \mathbf{f} und der $M \times N$ -Matrix \mathbf{A} .

In unserem Fall ist \mathbf{g} das Histogramm der *geschätzten* Gamma-Energien, für alle gemessenen Photonen, \mathbf{f} ist das Histogramm der wahren Gamma-Energien, \mathbf{A} die Migrationsmatrix des Energieschätzers und \mathbf{b} das aus den Off-Messungen abzuleitende Untergrundhistogramm. Ziel ist es nun einen Schätzer $\hat{\mathbf{f}}(\mathbf{g}, \mathbf{A}, \mathbf{b})$ zu finden.

Dies vernachlässigt den Effekt der Akzeptanz, Ereignisse, die nicht vom Teleskop gemessen wurden oder von der Analyse fälschlicherweise verworfen wurden. Die Akzeptanzkorrektur kann entweder zuletzt auf das geschätzte $\hat{\mathbf{f}}$ angewendet werden, oder die Matrix \mathbf{A} muss geeignet normiert werden.

Weitergehende Informationen finden Sie in [8] und [6, S. 241–254].

6.1 Naive SVD-Entfaltung

Ein einfacher Ansatz für die Inversion von (5) ist es, \mathbf{A} zu invertieren. Für nicht-quadratische Matrizen wird die Moore-Penrose-Pseudoinverse gebildet. Diese Lösung ist äquivalent zur Methode der kleinsten Quadrate.

$$\hat{\mathbf{f}} = \mathbf{A}^+(\mathbf{g} - \mathbf{b}) \quad (6)$$

6.2 Poisson-Likelihood-Entfaltung

Unter der Annahme, dass die Zählraten in \mathbf{g} Poisson-Verteilungen folgen, lässt sich ein Maximum-Likelihood-Fit durchführen. Die Wahrscheinlichkeit für einen Messwert g_i ergibt sich zu

$$P(g_i) = \mathcal{P}(g_i, \lambda_i) \quad (7)$$

mit

$$\boldsymbol{\lambda} = \mathbf{A} \cdot \mathbf{f} + \mathbf{b} \quad (8)$$

Die Likelihood ist dann

$$\mathcal{L} = \prod_{i=1}^M \mathcal{P}(g_i, \lambda_i) \quad (9)$$

numerisch ist es oft günstiger, die negative Log-Likelihood zu minimieren

$$-\ln \mathcal{L} = -\sum_{i=1}^M \ln \mathcal{P}(g_i, \lambda_i) \quad (10)$$

$$-\ln \mathcal{L} = \sum_{i=1}^M \ln(g_i!) - g_i \cdot \ln \lambda_i + \lambda_i \quad (11)$$

die relativ zu den Parametern konstanten Terme können weggelassen werden

$$-\ln \mathcal{L} = \sum_{i=1}^M -g_i \cdot \ln \lambda_i + \lambda_i \quad (12)$$

Der Schätzer für \mathbf{f} ergibt sich dadurch zu:

$$\hat{\mathbf{f}} = \operatorname{argmin}(-\ln \mathcal{L}(\mathbf{f} | \mathbf{A}, \mathbf{g}, \mathbf{b})) \quad (13)$$

6.3 Iterative Bayes Unfolding

Eine weitere, häufig in der Hochenergie- und Astroteilchenphysik genutzte Entfaltungstechnik ist das Iterative-Bayes-Unfolding. Hierzu siehe [8, S. 25].

6.4 Regularisierung

Aufgrund der schlechten Kondition der inversen Probleme, kommt es häufig zu unphysikalischen, stark schwankenden Ergebnissen. Um diese zu Unterdrücken, lassen sich alle Verfahren um Regularisierungsterme erweitern. Dies sind Strafterme, die unerwünschte Lösungen unterdrücken, zum Beispiel im Fall von Tikhonov-Regularisierung große zweite Ableitungen. [6] Hier muss jedoch beachtet werden, auf welche Größe die Regularisierung angewendet wird, da z. B. in der Gammaastronomie nur der Logarithmus des Flusses gegen den Logarithmus der Energie erwartbar eine kleine zweite Ableitung aufweist, nicht aber die Teilchenzählraten in geschätzter Energie.

7 Hinweise

7.1 Programmier-Hinweise

Es wird empfohlen, diesen Versuch mit Python zu bearbeiten. Die Dateien lassen sich leicht mit der Funktion `fact.io.read_h5py` aus dem Module `pyfact` (`pip install pyfact`) in `pandas.DataFrames` einlesen.

Zur Minimierung der negativen Likelihood kann `scipy.optimize.minimize` genutzt werden. Hier sollte beachtet werden, dass die Parameter von \hat{f} nur positiv sein dürfen.

Die Moore-Penrose-Pseudoinverse kann mit `numpy.linalg.pinv` berechnet werden.

Die Histogramme können mit `numpy.histogram` und `numpy.histogram2d` erstellt werden. Verwenden sie äquidistante Bins im Logarithmus der Energie, siehe `numpy.logspace`.

Eine Übersicht die auf fast alle für diesen Versuch benötigten Funktionen und Methoden eingeht, finden Sie unter https://github.com/maxnoe/scientific_python_notebooks/blob/master/smd_whirlwind_tour.ipynb

Literatur

- [1] M. Aartsen et al. „Neutrino emission from the direction of the blazar TXS 0506+056 prior to the IceCube-170922A alert“. In: *Science* 361.6398 (2018), S. 147–151. DOI: 10.1126/science.aat2890.
- [2] F. Aharonian et al. „The Crab nebula and pulsar between 500 GeV and 80 TeV: observations with the HEGRA stereoscopic air Cerenkov telescopes“. In: *The Astrophysical Journal* 614.2 (2004), S. 897.
- [3] J. Aleksić et al. „Measurement of the Crab Nebula spectrum over three decades in energy with the MAGIC telescopes“. In: *Journal of High Energy Astrophysics* 5 (2015), S. 30–38.
- [4] H. Anderhub et al. „Design and operation of FACT – the first G-APD Cherenkov telescope“. In: *JINST* 8.06 (2013), P06008.
- [5] A. Biland et al. „Calibration and performance of the photon sensor response of FACT — the first G-APD Cherenkov telescope“. In: *JINST* 9.10 (2014), P10012. URL: <http://stacks.iop.org/1748-0221/9/i=10/a=P10012>.
- [6] V. Blobel und E. Lohrmann. *Statistische und numerische Methoden der Datenanalyse*. Hamburg, 2012. URL: <http://www.desy.de/~blobel/ebuch.html>.
- [7] K. Brügge et al. *FACT-Tools v1.1.2*. 18. Dez. 2018. DOI: 10.5281/zenodo.2386762. URL: <https://github.com/fact-project/fact-tools>.
- [8] M. Bunse et al. „Unification of Deconvolution Algorithms for Cherenkov Astronomy“. In: *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE. 2018, S. 21–30. DOI: 10.1109/DSAA.2018.00012.
- [9] *FACT Open Data Homepage*. URL: <https://factdata.app.tu-dortmund.de>.

- [10] D. Heck et al. *CORSIKA: A Monte Carlo Code to Simulate Extensive Air Showers*, "Report FZKA 6019, Forschungszentrum Karlsruhe. Techn. Ber. 1998.
- [11] T.-P. Li und Y.-Q. Ma. „Analysis methods for results in gamma-ray astronomy“. In: *The Astrophysical Journal* 272 (1983), S. 317–324.
- [12] M. Nöthe und K. Brügge. *FACT Open Crab Sample Analysis*. 2019. URL: https://github.com/fact-project/open_crab_sample_analysis.