An airing of the Maximum Entropy/KKT method and its use for astrophysics

Uma discussão do método da Máxima Entropia/KKT e sua aplicação na astrofísica

Alexandre Humberto Andrei[1,2], Luciano Bedin[3] , Bruno Coelho[4], Leandro Guedes[5], Alexandre Lyra[1,6], Marcelo Mattos[7], Elias Rego[8]

[1]Observatório do Valongo, Universidade Federal do Rio de Janeiro
[2] Observatório Nacional, MCTIC, Rio de Janeiro
[3] Departamento de Matemática, Universidade Federal de Santa Catarina
[4] Instituto de Telecomunicações, Campus Universitário de Santiago, Aveiro, Portugal
[5] Fundação Planetário da Cidade do Rio de Janeiro
[6] Programa de Pós-Graduação em História das Ciências e das Técnicas e Epistemologia (HCTE), Universidade Federal do Rio de Janeiro
[7]Secretaria de Educação do Rio de Janeiro - SEEDUC
[8] Departamento de Matemática, Universidade Federal do Rio de Janeiro

oat1aha@gmail.com, bedin.luciano@gmail.com, brunodfcoelho@gmail.com, leandrolsguedes@gmail.com, alexandr@astro.ufrj.br, seacelo@hotmail.com, elias1211@gmail.com

Recebido: 27/04/2021 Aceito: 29/04/2021 Publicado: 29/04/2021

DOI:10.51919/revista_sh.v1i0.276

Abstract.*In this paper, after a brief history of the Maximum Entropy Principle (MaxEnt), we will discuss its generalization in an epistemological way with the use of the KKT methodology in discrete systems. For this purpose, we we will also make a brief history of the KKT method and its use. After this exhibition, we will proceed to the joint use of MaxEnt with KKT. This has already been done in some works found in the literature, but in the particular case of Astrophysics it is a new approach from our group, which we believe to be promising.*

Keywords. *Entropy. KKT Method. MaxEnt. Non linear programming.*

Resumo. Neste trabalho, após um pequeno histórico do Princípio da Máxima Entropia (MaxEnt), discutiremos de forma epistemológica sua generalização com a utilização da metodologia KKT em sistemas discretos. Para este objetivo faremos também um histórico resumido do método KKT e da sua utilização. Após esta exposição seguiremos para a utilização conjunta do MaxEnt com o KKT. Isto já foi feito em

alguns trabalhos da literatura, mas no caso particular da Astrofísica é uma abordagem nova do nosso grupo, a qual acreditamos ser promissora.

Palavras-chave. Entropia . Método KKT .MaxEnt . Programação Não-Linear.

1. Introdução

O Princípio da Máxima Entropia (MaxEnt) do físico americano Edwin Thompson Jaynes (1957) já foi alvo de trabalhos nossos em versões anteriores deste congresso (ANTUNES; LYRA, 2017). O ponto principal abordado foi sobre os fundamentos do MaxEnt, investigando o denominado ``Princípio da Razão Insuficiente''. Buscamos possíveis contribuições de G. Leibniz para estes fundamentos.

O trabalho atual surgiu a partir de uma ampliação da linha de pesquisa do grupo, após verificarmos que o Princípio da Máxima Entropia pode ser utilizado na Astrofísica juntamente com a metodologia KKT (KUHN; TUCKER, 1950). Detalharemos mais adiante esta metodologia, assim como o MaxEnt, visando o leitor que não conhece estesmétodos; após estes dois tópicos, discutiremos as questões principais do trabalho, que são referentes às aplicações na Astrofísica da metodologia conjunta MaxEnt/KKT. A discussão epistemológica e histórica está inserida em cada tópico abordado.

Em nossa abordagem partimos de algumas informações que o sistema fornece para estabelecer previsões mais gerais. A utilização do MaxEnt na Astrofísica é bastante promissora, uma vez que muitas vezes há poucos dados observacionais na região considerada e o princípio pode partirde poucos dados para fazer estimativas robustas. Em trabalho anterior (ANDREI et al., 2019) estabelecemos uma nova fórmula para a distribuição da função de luminosidade dos *quasars*a diferentes distâncias cosmológicas da nossa galáxia. Na investigação atual, estamos estimando a distribuição da *metalicidade* de estrelas na vizinhança solar, incluindo regiões com escassos dados observacionais (ANDREI et al., 2020).

O MaxEnt tem sido utilizado em áreas que vão dareconstrução de imagens tomográficas em Raios-X até previsões de distribuição de matéria escura nas galáxias. A bibliografia existente é enorme e, além disto, também encontramos na literatura discussões sobre a validade do MaxEnt(SHIMONY, 1985). Sabemos que estas discussões são importantes para a ciência, pois podem contribuir para se delinearem novos caminhos. Veja-se como a criação da Teoria da Relatividade de Einstein nasceu do questionamento do modelo Newtoniano e também do papel das transformações de Galileu no Eletromagnetismo *Maxwelliano*.

Um fato interessante é que com o nascimento da metodologia KKT, veio junto uma nova área muito importante para a ciência atual: a área de *Programação Não-Linear*. Abordaremos neste trabalho tais questões, revelando a grande importância da metodologia conjunta MaxEnt-KKTno ambiente da Astrofísica. Na seção 2 descreverEMOS o Princípio da Máxima Entropia; na seção 3, faremos o mesmo com o método KKT; já na seção 4, abordaremos a interação destes dois métodos e sua aplicação à Astrofísica; na última seção deixamos algumas conclusões do trabalho.

2. O Princípio da Máxima Entropia

O conceito de entropia foi introduzido na Ciência no século XIX nos trabalhos de Clausius. Nos primórdios destes estudos se encontrava a Termodinâmica, a qual trouxe uma grande contribuição para o mundo moderno. Os trabalhos de Maxwell, de Boltzmann, e posteriormente a contribuição de Gibbs e o estabelecimento da Mecânica Estatística foram fundamentais para a ciência moderna. Aplicações destes conhecimentos vão desde questões que envolvem o estudo do Clima, da Cosmologia, e até as técnicas da Engenharia. Várias áreas se beneficiaram com estas investigações e encontramos vasta bibliografia. Segundo Jaynes (1978, p.14), Boltzmann abriu o caminho para o Princípio da Máxima Entropia, colocando questões como: (a) de quantas maneiras diferentes um determinado número de moléculas pode ser distribuído? (b) entre todas as distribuições possíveis, qual é a mais provável? A resposta de Boltzmann serviu como seu ponto de partida. A distribuição mais provável é aquela que pode ser realizada pelo maior número de caminhos possíveis, isto é, aquela que pode ser *maximizada sujeita a certas restrições*.

O matemático e engenheiro norte-americano Claude E. Shannon(1916-2001) publicou (1948), a sua Teoria Matemática da Informação. A teoria mostrou que a maioria dos canais de comunicação possui uma capacidade ou uma taxa de transmissão e que a informação só pode ser transmitida através do canal se, e somente se, a quantidade de informação enviada pela fonte do canal não exceder a sua capacidade de transmissão. Por esse motivo, Shannon procurou otimizar os meios de comunicação daquela época, diagnosticando a capacidade de transmissão de cada canal e o nível de confiabilidade em cada informação enviada, desde o emissor até o receptor. De acordo com Shannon (1948, p.11),toda informação está associada à incerteza e a medida utilizada para quantificar esta incerteza tem a mesma expressão matemática da entropia na Mecânica Estatística, exceto pela presença de uma constante K, positiva. Apesar da diversidade de problemas que puderam ser tratados pela Teoria da Informação, ainda não existia sua ligação com a Mecânica Estatística, pois o conceito de entropia estava vinculado somente aos problemas da Física. O mero fato de que a mesma expressão matemática $H = -K\sum p_i \operatorname{Log} p_i$, ocorrer tanto na Física quanto na Teoria da Informação não estabelece, por si só, qualquer ligação entre esses campos. Porém, Jaynes em 1957 interpretou pela Mecânica Estatística a entropia da teoria de Shannon: a *entropia de uma distribuição de probabilidade* p_i. Desta forma criou o seu *Princípio da Máxima Entropia*, afirmando que "a teoria da informação forneceu um critério para atribuir distribuições de probabilidade com base num conhecimento parcial, que leva a um tipo de inferência estatística denominada estimativa de máxima entropia." (JAYNES, 1957, p. 620). O conceito de entropia, que anteriormente estava vinculado somente aos problemas da Física, pôde ser generalizado com a teoria do Shannon. Nesta nova teoria, para fazer inferências com base em informações parciais, devemos usar a distribuição de probabilidade que tenha a máxima entropia H, a partir de qualquer informação conhecida. Pode-se dizer que se trata de predizer aquilo que não se vê. Assim, a expressão da entropia pode ter um significado bastante independente da Termodinâmica.

"O grande avanço que a teoria da informação forneceu está na descoberta de que um único critério disponível para "quantificar a incerteza" é dado por uma distribuição de probabilidade discreta que esteja de acordo com alguma informação disponível." **(JAYNES, 1957, p. 621)**

Reconhecendo que na análise de qualquer sistema, a incerteza pode ser decorrente da falta de informação ou de informações incompletas. Para cada quantidade de informações que o sistema fornece temos também os vínculos (ou *restrições*) da distribuição p_i. Quanto maior é a quantidade de informações, menor é a entropia (ou a incerteza) do sistema. Quando não temos nenhuma informação disponível a incerteza (ou entropia), é máxima e o sistema é o mais aleatório possível. Segundo Jaynes, "*uma ampla distribuição de probabilidade representa mais incerteza do que uma precisão acentuada*" (JAYNES, 1968). Agora, na nova teoria, incerteza e entropia são sinônimos (JAYNES, 1957, p.622), sua medida está associada à distribuição de probabilidade.

Concisamente podemos dizer de que o Princípio de Máxima Entropia possibilita tratar vários fenômenos que envolvem imprevisibilidade e incerteza, como ocorre frequentemente na análise de sistemas estatísticos, na mecânica quântica e na termodinâmica. Veremos mais adiante a formulação matemática do Princípio.

3. O Método KKT

Ainda que na historiografia da metodologia KKT hajam vários personagens envolvidos, escolhemos apenas alguns deles para aqui mencionarmos. Os interessados em um maior aprofundamento do tema podem verificar nas referências citadas, dentre elas (KJELDSEN, 2000). Hoje em dia este método, foi denominado KKT, devido aos pesquisadores envolvidos, Karush-Kuhn-Tucker.

3.1. O Trabalho Precursor de Karush

Em dezembro de 1939, o norte-americano William Karush(1917 - 1997)recebeu um diploma de mestre em matemática daUniversidade de Chicago. Foi professor de matemática da "*California State University*" em Northridge; muito tempo depois ficou conhecido por sua contribuição ao método KKT.A sua dissertação de mestrado de 1939, teve o título "Mínimos de funções de várias variáveis com desigualdades como condições colaterais", sendo o primeiro desenvolver essas condições do futuro método KKT. Na realidade tornou-se conhecido décadas após sua tese, em um artigo de conferência realizado por Harold W. Kuhn e Albert W. Tucker. Como físico trabalhou para o Projeto Manhattan, embora tenha assinado depois a conhecida petição Szilárd e se tornado depois disso um ativista pela paz. Hoje diríamos quetal problema de otimização sujeito a restrições de desigualdade pertence ao domínio da*Programação Não-linear*, mas esta denominação somente surgiu bem posteriormente. O projeto de Karushfoi proposto por seu orientador, Lawrence M. Graves (KJELDSEN, 2000). Na introdução da tese, Karush declarou o propósito de seu trabalho e também deu sugestões de onde buscar a motivação por trás da proposta do problema. Propôs determinar as condições necessárias e suficientes para um mínimo relativo deuma função f $(x_1, ..., x_n)$que satisfaz as equações g_a (x) = 0 (a = 1, ..., m), onde as funções f e g_a têm derivadas contínuas até pelo menos a segunda ordem. Também retomao correspondenteproblema satisfazendo as desigualdadesg_a (x) $\geq$0 (a = 1,, 2 ..., m),onde m pode ser menor, igual ou maior que n. Ao apresentar a tese em dezembro de 1939, nadaaconteceu, o trabalho não foi publicado, ninguém o incentivou a divulgar seu resultado, eaparentemente considerou-se que não era interessante.

3.2. O Trabalho de Kuhn, Tucker e John

Albert William**Tucker**(1905-1995)**,** matemático de Princeton, nasceu no Canadá, e recebeu o diploma de bacharel em matemática pela Universidade de Toronto em 1928.

Harold W. **Kuhn**(1925-2014), 20 anos mais jovem que Tucker, nasceu na Califórnia, e fez o bacharelado em ciências pelo *California Instituteof Technology* em 1947, e então se mudou para Princeton, onde escreveu sua tese de doutorado em Teoria de Grupos. No verão de 1950, no Segundo Simpósio de Berkeley, na Califórnia, sobre Estatística Matemática eProbabilidade,Tucker,que geralmente era conhecido como topólogo, deu uma palestra baseada em um trabalho conjunto com o jovem matemáticoKuhn, quetinha acabado de concluir seu doutorado em Princeton. As palestras foram publicadas pela conferência, e pela primeira vez apareceu o nome "Programação Não-Linear", que foi o título que Kuhn e Tucker escolheram para seu artigo (KUHN e TUCKER, 1950, 1951). No artigo, Kuhn e Tucker provaram o teorema principal da teoria. Este teoremadá as condições necessárias para a existênciade uma solução ótima para um problema de programação não-linear. O resultado é famoso e, não muito depois de sua publicação, tornou-se conhecido como oTeorema de Kuhn-Tucker. Apesar disto, Kuhn e Tucker, como vimos, não foram os primeiros matemáticos que provaram o teorema. Em livros modernos sobre programação não-linear, costuma haver notas dizendo que Karush provouo teorema em 1939, em sua tese de mestrado da Universidade de Chicago. Hoje em dia é citado como o "teorema de Karush-Kuhn-Tucker", ou KKT, reconhecendo o trabalho de Karush.Além disto, Fritz John (1910-1994), um matemático alemão, obteve um resultado semelhante no seu trabalho de 1948, na coleção de ensaios para o 60º aniversário do seu orientador Richard Courant. Este trabalho de 1948foi inicialmente rejeitadopelo Duke MathematicalJournal. Mas o artigo saiu dois anos antes do trabalho de Kuhn e Tucker, e novamente ninguém percebeu. Hoje ele é provavelmente mais conhecido por seu trabalho em equações diferenciais parciais, mas também fez contribuições importantes nos campos da geometria, análise e elasticidade. O interessante é que apenas dois anos depois, quando Kuhn e Tucker derivaram o mesmo resultado, este se tornou famoso quaseinstantaneamente e assim foi lançada a nova área de pesquisa matemática, a *Programação Não-Linear*.

Por causa destes diferentes trabalhos, abordando o mesmo problema, discute-se (KJELDSEN, 2000) se o teorema de Kuhn-Tucker pode ter sido uma descoberta múltipla de Karush, John, Kuhn e Tucker. É interessante ainda como o conteúdo do teorema foi recebidode formas diferentes pela comunidade matemática.Várias reflexões são feitas a partir destes fatos. Será que este foi realmente o mesmo resultado que eles conseguiram? Será que foi uma descoberta múltipla? Por que as reações do meio científico foram tãodiferentes nos três casos? Por que nada aconteceu nas duas primeiras vezes? Por que o trabalho de Kuhn e Tucker teve um impacto tão grande? Perguntas como estas são discutidas em (KJELDSEN, 2000), onde encontramos ainda que em 1975, Kuhn afirmou para Karush em carta sobre o seu trabalho, ``*Primeiro, deixe-me dizer que você tem prioridade clara nos resultados conhecidos como as condições de Kuhn-Tucker (incluindo a qualificação de restrição). Pretendo deixar as coisas o mais claras que puder em minha palestra..."* Kuhn estava se referindo a uma palestra que lhe pediram para fazer sobre a história da Programação Não-linear. Finalmente, podemos sintetizar, com a definição de F. John, sobre a nova metodologia: "*O método KKT trata de uma extensão da regra do multiplicador de Lagrange para o caso onde as condições são desigualdades em vez de equações*" (KNELDSEN, 2000).

3.3. Programação Não-Linear

Vimos que o teorema KKT se estabeleceu como método de Programação Não-linear, ele nos interessa, em especial, quanto à solução de problemas astrofísicos balizados por condições de desigualdade. Mas cabe trazer, de modo breve, a circunstância histórica mais estendida e o que dela emerge quanto ao enfoque especificamente epistemológico.

O problema paradigmático da programação não-lineaté o isoperimétrico, isto é, a maximização da área delimitada por um perímetro, cuja solução foi obtida na Grécia antiga, atribuída a Zenodoro, e que, no entanto, foi demonstrada apenas 1800 anos depois por Johann Bernoulli. À época, no rastro da álgebra islâmica, do renascentismo científico e da revolução astronômica de Copérnico, Galileu e Kepler, houve como que uma disputa entre a plêiade de matemáticos europeus, de tal sorte que diferentes provas foram quase simultaneamente obtidas por Newton, Leibniz e pelo próprio irmão de Bernoulli, Jakob. Ou seja, na voga do desenvolvimento do cálculo infinitesimal, diferencial e dos métodos de aproximação geométricos, por séries e polinômios. O cálculo diferencial permitiu um tratamento satisfatório dos problemas de otimização e inequações, em regra não-lineares. Para chegar ao KKT, bastao título da tesede W. Karush, “Minima of functions of several variables withinequalities as sideconditions".

O outro ponto que vale ser salientado é como o KKT, ou métodos de otimização, de busca de máximos e mínimos, de condições por inequalidades, em suma problemas não-lineares, trazem a embocadura da discussão a respeito da fronteira entre a filosofia e a epistemologia. Certo que a epistemologia utiliza a construção lógica estruturante da filosofia, enquanto a filosofia aplicada busca o empiricismo como verificação da razão pura. Estas disciplinas se entrelaçam e complementam, a epistemologia apresenta-se de bom grado como ramo da filosofia. Uma estuda a verdade, a outra o conhecimento. Na era moderna, a epistemologia se defronta com a questão da ciência como senhora ou serva dos afazeres humanos. Não raro o mesmo discurso demanda que a sociedade siga o ditame da ciência e de seus acólitos, cientistas e técnicos, ao mesmo tempo que reclama um controle moral e social de seus rumos, metas e práticas. De modo vívido, podemos ecoar Bertrand Russel, que tratou Aristóteles como uma das maiores infelicidades da humanidade, por trazer à sua cuidadosa descrição da natureza, fundações de simetria, ou seja, abstratas. Ou preferir as críticas de Sartre à pesquisa nuclear, ou de Foucault àquilo que hoje chamamos de neurociência. O assunto é extenso e de debate aberto. Aqui nos interessa apontar que o desenvolvimento de abordagens não-lineares, herdeiras como vimos das técnicas de aproximações e do cálculo diferencial, visando maximização e otimização, encerra o reconhecimento de que a natureza não contém absolutos, nem mesmo imperativos, exceto como representações locais ou hipóteses simplificadoras.

4. O Método da Máxima Entropia/KKT em Astrofísica

A metodologia do KKT introduz uma possibilidade muito importante para a Astrofísica, contudo poucos trabalhos conhecemos com esta sistemática em todo campo da Astronomia. Alguns exemplos são: em Canuet al. (2016), onde encontramos um caso de KKT aplicado à Astronomia, porém sem MaxEnt; e em (THIÉBAUT; YOUNG (2017) sobre reconstrução de imagens astronômicas, onde os dois métodos são citados.

Já a metodologia do Princípio da Máxima Entropia pode ser encontrada facilmente na literatura, como no trabalho de (ANDREI et al., 2019). Neste trabalho, partimos de um conjunto de dados, dos quais temos uma certa função $f(x_i)$ onde $x_i = (x_1, x_2, \ldots, x_n)$ é um conjunto de valores discretos, além disto temos observáveis físicos $X_k$cujos valores médios de cada um deles é dado por $\langle X_k \rangle = \sum_i p_i X_{ki}$, mas não temos as probabilidades

$p_i(s)$. O princípio utiliza para obter as probabilidades p_i a definição de entropia $H = -\sum_i p_i \log p_i$. A partir disto se aplica a metodologia dos multiplicadores de Lagrange para maximizar a função H sujeita aos vínculos (ou restrições) do sistema, que são, a normalização de p_i ou seja, $\sum_i p_i=1$, e os valores médios dos observáveis. Isto é o básico do MaxEnt, mais detalhes estão nas referências citadas. Não poderíamos deixar de mencionar que o MaxEnt também obtém, na sua formulação, a probabilidade uniforme no caso de sistemas dos quais não tenhamos nenhum valor médio de observáveis. Aí obtemos $p_i = 1/N$, onde N é o número de estados possíveis, a entropia aí é máxima.

A maior diferença do MaxEnt com o método KKT, é que com este podemos assumir que os observáveis X_k podem ser fornecidos através de desigualdades, ou inequações. Por exemplo, que $X_k \leq A$, ou seja, $X_k - A \leq 0$, em outras palavras, podemos assumir que certos observáveis físicos pertencem a um intervalo de variação. Isto é muito importante na Astrofísica, ampliando a aplicabilidade do MaxEnt. Os dados observacionais podem ser escassos em certas áreas observadas, contudo suficientes para estabelecer que tenham um valor mínimo ou máximo.

No formalismo matemático, se assume também a normalização da distribuição de probabilidade p_i , significando, ressaltemos, que nenhum evento tem possibilidade infinita de realizar-se. Então, utilizando os multiplicadores de Lagrange, pelo KKT, encontramos os extremos da distribuição, através uma Lagrangeana, que dá a expressão matemática que permite maximizar a entropia, e finalmente obter a distribuição de probabilidade. Dependendo do problema abordado e do número de valores médios assumidos, ou de desigualdades iniciais, as distribuições de probabilidade admitem uma representação exponencial, algo da forma $p_k= \exp(- \lambda_1 - \lambda_2 X_k)$, onde λ_1 e λ_2 são multiplicadores de Lagrange e X_k é um observável. O significado físico é que a realidade é probabilística, nenhum evento válido jamais tendo possibilidade quer total, quer nula de se verificar. Os multiplicadores são fornecidos pelos dados do problema, assim como os valores da desigualdade. Esta metodologia é detalhada em outro trabalho do nosso grupo (ANDREI et al., 2021), em preparação.

5. Conclusões

É interessante mencionar a partir da história do KKT, como um trabalho científico importante fica esquecido e é, de forma inusitada, décadas após o trabalho inicial, “”redescoberto” pela comunidade científica. Sabemos que isto já aconteceu em outras áreas científicas, o que levanta a questão da neutralidade científica, perante grupos de interesse. Além disto, discute-se na literatura se foi ou não uma descoberta múltipla.
Enfatizamos aqui que na Astrofísica a metodologia MaxEnt/KKT pode ser mais aplicada, pois os problemas astrofísicos são muito propensos a terem observáveis com valores médios apenas delimitados em certas regiões, onde existem poucos dados observacionais. A metodologia MaxEnt/KKT fornece uma teoria para tratar estes problemas. Conforme dissemos, investigamos atualmente a *metalicidade* de estrelas da vizinhança solar com esta metodologia. Outras questões relacionadas e apontadas neste trabalho poderão ser objetos de futuros desenvolvimentos. Gostaríamos de estimular outras pesquisas com esta metodologia. Finalizamos com JAYNES (1990, p.1):
...”o fato de que uma certa distribuição de probabilidade maximize a entropia sujeita a certas restrições que representam nossa informação incompleta, é a propriedade fundamental que justifica o uso dessa distribuição para inferência; concorda com tudo o que é conhecido, mas evita cuidadosamente assumir

qualquer coisa que não é conhecida. É uma transcrição para a matemática de um antigo princípio de sabedoria; e realiza automaticamente a síntese necessária dos pontos de vista de Gibbs e Jeffreys."

Agradecimentos

A.H.A. agradece a Bolsa PQ CNPq #302870/2017-2. B.C. agradece a Bolsa POCI-01-0145- FEDER022217 subvencionada por COMPETE2020 e FCT.

Referências Bibliográficas

ANDREI, A. et al. 2019, The Principle of Maximum Entropy and the Luminosity Function of Quasars, **Monthly Notices of the Royal Astronomical Society**, v.488, p.183-190.

ANTUNES, M. e LYRA, A., O Princípio da Máxima Entropia e o Problema da Razão Insuficiente, **Anais do Scientiarum História XI**, Rio de Janeiro: UFRJ,2018.

CANU *et al.* **Introduction to optimization with applications in astronomy and astrophysics,** em :https://hal.archives-ouvertes.fr/hal-01346134.

JAYNES, E. T, (1957) **Phys. Rev**. 106, p.620;``Prior Probability", IEEE Transations On Systems Science and Cybernetics, vol.4 sec.4 N.3 (1968); "Where do we Stand on Maximum Entropy?, Maximum Entropy Formalism Conference, Massachusetts Institute of Technology, May 2-4, (1978); "Notes on Present Status and Future Prospects." In Maximum Entropy and Bayesian Methods, edited by W. T. Grandy and L. H. Schick. Kluwer, Springer, Wyoming, USA (1990).

KUHN, H.W. AND TUCKER, A.W., 1950. **Nonlinear programming, In Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability**, J. Neyman, Ed., pp. 481–492. Berkeley; Berkeley Symp. on Math. Statist. and Prob.Proc. Second Berkeley Symp. on Math. Statist. and Prob. (Univ. of Calif. Press, 1951), 481-492 , Nonlinear Programming.

KUHN, HAROLD W. **Nonlinear programming:** a historical view, Nonlinear programming (Proc. Sympos., New York, 1975) Amer. Math. Soc., Providence, R. I., 1976, pp. 1–26. SIAM-AMS Proc., Vol. IX. MR 0403674,

KARUSH, W. , 1939. **Minima of Functions of Several Variables with Inequalities as Side Conditions** . Dissertation, Department of Mathematics, University of Chicago. Illinois;

KJELDSEN, T. H. **A Contextualized Historical Analysis of the Kuhn–Tucker Theoremin Nonlinear Programming:** The Impact of World War II, In: Historia Mathematica 27 (2000), 331–361.

SHANNON, C.E. “**The mathematical Theory of Communication**“, Bell System Tech. J.27, 379, 623 (1948).

SHIMONY, A. The Status of the Principle of Maximum Entropy, **Synthesis**, v.63, p-35-53 (1985).

THIÉBAUT et YOUNG (**Journ. of the Optical Soc. America A**, v34, n.6, p.904 (2017)).