e Wienner [2] vem idéia de discutir informação do ponto de vista de volume, quantidade com o objetivo de entender limitações para a transmissão em canais de comunicação. Isto inclui como codificar uma mensagem, comprimindo-a e como, no outro lado do canal de comunicação, reobté-la. Não foi discutida a qualidade da informação ou a utilidade da informação: se fosse, toda a teoria não teria utilidade para descrever por exemplo meios como a televisão.... Isto não é totalmente piada. Se olharmos uma máquina de processamento de informação, como por exemplo um sistema nervoso de um animal, o conceito do valor da informação toma uma posição muito mais central. Há vantagens evolutivas em realizar inferência de uma forma frente a outra. Mais sobre tudo isso em aulas posteriores [3].

[2] Aparentemente Alan Turing também teve estas idéias durante a 2da Guerra Mundial enquanto tentava decifrar códigos, mas esse trabalho teve menos influência porque era confidencial. Ver livro de Good.

[3] Não cabe no curso introdutório de Mecânica Estatística

6.1.1 *Um método variacional*

O caminho para as aplicações está aberto. O método de escolha das distribuições de probabilidade é de maximizar a ignorância. De todas as distribuições compatíveis com os vínculos, escolhemos a que introduz a menor quantidade de informação adicional aos vínculos. Reduzimos a inferência a um método variacional. Isto não deve vir como uma surpresa, pois toda a Física é essencialmente redutível a métodos variacionais. A pergunta profunda é se esses métodos estão relacionados ou se são independentes. A reposta que está emergindo é que todos esses métodos estão relacionados. E isto transcende a Física e se estende por outras áreas. Uma contribuição importante de Jaynes foi a percepção que a maximização da entropia sujeita aos vínculos da informação leva a um método fundamental de inferência, chamado por ele de MaxEnt. Outros nomes estão associados a extensão de idéias de entropia como método de inferência fundamental [4]. Em particular tem se conseguido nos últimos anos uma formulação muito mais satisfatória e menos ad hoc dos princípios por trás deste método.

[4] ver A.C

A generalização do método de máxima entropia para inferência em problemas com variáveis em variedades contínuas pode levar a problemas conceituais que serão atacados a seguir.

6.2 *Entropia Relativa*

As deduções aqui deveriam ser lidas após treinamento sobre métodos variacionais, que será abordado nos próximos capítulos. Portanto talvez seja recomendável pular esta seção numa primeira leitura. A entropia de Shannon vai ser usada para inferências em quase todo o resto do livro. Mas esta não é a última palavra em mecanismos de inferência. A utilidade de entropia de Shannon foi a de *atribuição* de probabilidades. Isso deve dar a impressão que do nada, as probabilidades são construidas através da imposição de certos vínculos sobre funções dos graus de liberdade. Dificilmente a inferência é feita a partir de nenhum conhecimento prévio. Para qualquer situação é mais razoável que um estado de crenças seja modificado pela imposição de novos vínculos. Isso nos leva à definir Informação como tudo aquilo que leva a modificar as crenças codificadas em distribuições de probabilidades. Procuramos um mecanismo que sirva para *atualizar* as crenças a partir de uma crença *a priori*. A diferença importante pode ser vista no confronto de *atualizar* a partir de um estado de conhecimento *apriori* versus *atribuir* a partir do vácuo informacional.

Qual deve ser o mecanismo de mudar crenças? Há em geral muitas distribuições de probabilidade que satisfazem um certo conjunto de vínculos. A idéia é fazer um ranking de todas elas. Isto significa atribuir um número a cada distribuição de probabilidades e escolher aquela que maximiza o ranking de preferências. Ou seja devemos construir um funcional que atribui um número a cada distribuição. Claro que esse número deve depender no estado de crença antes da imposição dos novos vínculos. Como escolher o funcional? Novamente procuramos casos simples onde sabemos algo. Impomos como desejo que o funcional dê a resposta esperada neles. O método de obtenção do funcional é o de eliminação de todos os funcionais que falham em satisfazer esses casos simples. Resta algum funcional após este processo? Poderia ser que não, mas não é o caso. Chamaremos este funcional de Entropia Relativa $S[p||q]$, entre a distribuição p candidata a ser escolhida pelo processode inferêcnia e a distribuição *a priori* q. Também pode ser chamado de (menos) a divergência de Kullback-Leibler (KL). Por economia, será simplesmente chamada entropia. A menos de pequenos detalhes, a dedução segue A. Caticha de forma simplificada.

6.2.1 *Desiderata Entrópica*

Os desejos expostos a seguir são considerados em adição aos do Capítulo 1. Portanto a teoria de probabilidades é a estrutura matemática adequada. Mas agora faremos algumas demandas adicionais. Pedimos ao leitor que encontre argumentos contra os seguintes desejos para uma teoria de infêrencia:

- DE_1: **Generalidade**: O método deve ser Universal.

Queremos lidar com informação incompleta e o algoritmo de procedimento deve ser o mesmo independente de qual o tipo de problema de inferência que estamos tratando. A descrição do problema, através da escolha dos graus de liberdade apropriados e dos vínculos impostos levará em conta a natureza explícita do problema.

- DE_2: **Parcimônia** A atualização deve ser minimalista.

De todas as distribuições $p(x)$ que satisfazem o vínculo devemos escolher a que menos mude a distribuição *a priori* $q(x)$. Claro que o que quer dizer mínimo deve ser definido. Mas há um caso simples em que mínimo é fácil de decidir. No caso em que não há nova informação, então a menor mudança é não fazer nada. Nosso método deve ser tal que $p(x) = q(x)$ após a incorporação de nenhuma informação. Parece muito trivial mas será útil.

- DE_3: **Localidade** no espaço de configurações.

Quando o espaço de configurações pode ser dividido em duas partes disjuntas, informação que diz respeito somente a uma das partes, ao ser incorporada, não deve alterar a atribuição de probabilidades às configurações da outra.

- DE_4: **Invariância**: A escolha dos rótulos dos graus de liberdade é uma convenção: não deve alterar o resultado da Inferência.

Para sistemas de variáveis que tomam valores num sub espaço de $\mathbb{R}^N$, a escolha do sistema de coordenadas não deve interferir no resultado do problema. Assim o funcional deve ser invariante ante mudanças (contínuas diferenciáveis) do sistema de coordenadas.

- DE_5: **Independência**: Sistemas independentes devem ser descritos da mesma forma quando estudados separadamente ou em conjunto.

Há situações em que sabemos que há sistemas independentes. Consideremos uma garrafa de café e a galáxia $X9$. Ao fazer predições sobre a termodinâmica do café podemos (i) tratar o sistema de café sozinho, ou (ii) incluir os graus de liberdade da galáxia $X9$ além dos do café, e marginalizar a distribuição, integrando sobre as configurações da galáxia. Os dois métodos devem dar o mesmo resultado sob pena de ao realizar medidas experimentais, poder distinguir entre os casos em que a galáxia foi ou não incluida. Note que poderiamos, caso dessem diferentes, fazer astronomia olhando para uma garrafa de café.

6.2.2 DE_3 : *Localidade*

Começamos com as consequências de DE_3. As configurações de um sistema formam o conjunto $\chi = \{x_1, x_2, ...x_n\}$. Podemos separar em dois conjuntos $i \in A$ e $j \in B$, tal que $A \wedge B = \emptyset$ e $A \vee B = \chi$, ou seja mutuamente exclusivos e exaustivos. Um caso simples é quando a informação é dada através dos seguintes vínculos

$$P_A = \sum_{x_i \in A} p_i, \tag{6.12}$$

$$P_B = \sum_{x_j \in B} p_j, \tag{6.13}$$

$$\text{com} \quad P_A + P_B = 1, \tag{6.14}$$

$$G = \sum_{x_j \in B} p_j g(x_j), \tag{6.15}$$

onde $p_k = p(x_k)$. Maximizar a entropia sujeita a estes vínculos leva a

$$0 = \delta\left[S - \lambda_1\left(\sum_A p(x_i) - P_A\right) - \lambda_2\left(\sum_B p(x_j) - P_B\right) - \lambda_3\left(\sum_B p(x_j)g(x_j) - G\right)\right]$$

para variações independentes de quaisquer dos p. Temos n equações:

$$\frac{\partial S}{\partial p_i} = \lambda_1 \quad (6.16)$$

$$\frac{\partial S}{\partial p_j} = \lambda_2 + \lambda_3 g(x_j). \quad (6.17)$$

$$(6.18)$$

Em geral teriamos, para $k \in A$ ou $k \in B$, chamando $\frac{\partial S}{\partial p_k} = f_k$, a depêndencia

$$\frac{\partial S}{\partial p_k} = f_k(p_1.....p_n, q_1....q_n). \quad (6.19)$$

Comparando as equações 6.16 e 6.19 vemos que

$$\lambda_1 = f_i(p_1.....p_n, q_1....q_n) \quad (6.20)$$

$$= f_i(p_i, q_i). \quad (6.21)$$

A segunda equação decorre de que mudando a informação que altera p_j do conjunto B, poderiamos mudar o lado direito sem alterar o multiplicador de Lagrange que é constante. Ainda mais, podemos considerar todas as possíveis divisões de χ em subconjuntos exaustivos e mutuamente exclusivos. Eliminariamos a dependência em todos os indices diferentes de i. Assim

$$\frac{\partial S}{\partial p_k} = f_k(p_k, q_k) \quad (6.22)$$

Integrando, vemos que a forma geral do funcional é

$$S = \sum F_k(p_k, q_k)$$

onde F_k com $f_k = \frac{\partial F_k}{\partial p_k}$ são funções ainda desconhecidas. A generalização para variáveis que tomam valores reais:

$$S = \int F(p(x), q(x), x)dx. \quad (6.23)$$

6.2.3 *DE_4: Invariância*

Vamos começar com um problema em uma dimensão. Uma densidade de probabilidades de uma variável X, $m(x)$ satisfaz

$$\int m(x)dx = 1$$

Podemos mudar variáveis $x \to x'$ e a densidade mudará $m(x) \to m'(x')$. A probabilidade de uma região dx em x deve ser preservada, portanto

$$m(x)dx = m'(x)dx'$$

e isso deve valer para a transformação de qualquer densidade. Em mais de uma dimensão teremos

$$m(x) = m'(x')J(x')$$

onde $J(x') = \det|\frac{\partial x'}{\partial x}|$ é o Jacobiano da transformação $x \to x'$.

A estratégia a seguir é a seguinte: fazemos a inferência no sistema de coordenadas x, obtendo um resultado. A seguir, a fazemos no referencial x'. Transformamos os x' de volta a x e comparamos. Os dois resultados deveriam dar igual. Mas não ocorre a não ser que imponhamos condições sobre a função F que aparece na equação 6.23. Isso restringe ainda mais os funcionais sobreviventes. Antes de prosseguir, facilita introduzir uma densidade $m(x)$ que por agora está à nossa disposição escolher. A motivação é que a razão de duas densidades tem propriedades simples de transformação

$$\frac{p(x)}{m(x)} = \frac{p'(x')}{m'(x')}, \tag{6.24}$$

pois o Jacobiano se cancela [5]. Assim consideramos que a equação 6.23 pode ser escrita

[5] Cabe ressaltar que deveriamos ter cuidado de afirmar que as regiões em que as diferentes distribuições se anulam são as mesmas e são excluidas destas considerações.

$$\begin{aligned} S &= \int F(p(x),q(x),x)dx = \int \frac{1}{m(x)}F(\frac{p(x)}{m(x)}m(x),\frac{q(x)}{m(x)}m(x),x)m(x)dx, \\ &= \int \Phi(\frac{p(x)}{m(x)},\frac{q(x)}{m(x)},m(x),x)m(x)dx, \end{aligned} \tag{6.25}$$

onde introduzimos a ainda desconhecida função

$$\Phi(u,v,m,x) = \frac{1}{m}F(um,vm,x).$$

Queremos mudar as coordenadas de forma geral, mas é conveniente se restringir a um caso fácil. O vínculo imposto será

$$\int p(x)a(x) = A,$$

onde a função $a(x)$ é um escalar, o que significa que $a(x) \to a'(x') = a(x)$, não muda ante transformações de coordenadas. A normalização é um exemplo de vínculo escalar, portanto não é preciso considerá-lo separadamente. O cálculo variacional $\delta\,(S\ +\text{vínculos}\,) = 0$, leva a

$$\begin{aligned} 0 &= \frac{\delta}{\delta p(x)}\left(S - \lambda\int p(x)a(x)dx\right), \\ &= \dot{\Phi}(\frac{p(x)}{m(x)},\frac{q(x)}{m(x)},m(x),x) - \lambda a(x), \end{aligned} \tag{6.26}$$

onde

$$\dot{\Phi}(u,v,m,x) = \frac{d}{du}\Phi(u,v,m,x).$$

Analogamente, começando no sistema de coordenadas x', chegamos a

$$0 = \dot{\Phi}(\frac{p'(x')}{m'(x')}, \frac{q'(x')}{m'(x')}, m(x'), x') - \lambda' a'(x'). \quad (6.27)$$

Usando a equação 6.24, esta última equação pode ser reescrita como

$$0 = \dot{\Phi}(\frac{p(x)}{m(x)}, \frac{q(x)}{m(x)}, m'(x'), x') - \lambda' a'(x'). \quad (6.28)$$

Para vínculos escalares, segue que, a razão

$$\frac{\dot{\Phi}(\frac{p(x)}{m(x)}, \frac{q(x)}{m(x)}, m(x), x)}{\dot{\Phi}(\frac{p(x)}{m(x)}, \frac{q(x)}{m(x)}, m'(x'), x')} = \frac{\lambda}{\lambda'} \quad (6.29)$$

é uma constante, pois os multiplicadores de Lagrange não dependem de x nem x'. Como isto vale para qualquer transformação contínua e diferenciável de $x' = \phi(x)$, podemos olhar para diferentes escolhas de ϕ e eliminar candidatos. Primeiro, olhamos para transformações com Jacobiano igual a 1 em todo o espaço. Segue que $m(x) = m'(x')$, portanto a única diferença entre o numerador e o denominador é no quarto argumento. Como a razão é constante, $\dot{\Phi}$ não pode depender do quarto argumento x ou x'. Agora consideramos, novamente com Jacobiano igual a 1, a função $\phi(x) = x, x \in \bar{D}$ e $\phi(x) \neq x, x \in D$. Se $x \in \bar{D}$, a razão é 1:

$$\begin{aligned} 1 &= \frac{\dot{\Phi}(\frac{p(x)}{m(x)}, \frac{q(x)}{m(x)}, m(x))}{\dot{\Phi}(\frac{p(x)}{m(x)}, \frac{q(x)}{m(x)}, m'(x'))} \\ &= \frac{\lambda}{\lambda'} \end{aligned}$$

Mas para $x \in D$, não muda a razão. Portanto $\dot{\Phi}$ não pode depender agora do seu terceiro argumento e

$$1 = \frac{\dot{\Phi}(\frac{p(x)}{m(x)}, \frac{q(x)}{m(x)})}{\dot{\Phi}(\frac{p(x)}{m(x)}, \frac{q(x)}{m(x)})} \quad (6.30)$$

é trivialmente satisfeita. Segue que o funcional só pode ter esta estrutura:

$$S = \int \Phi(\frac{p(x)}{m(x)}, \frac{q(x)}{m(x)}) m(x) dx. \quad (6.31)$$

6.2.4 DE_2 : *Parcimônia*

Temos à nossa disposição este intruso $m(x)$. Podemos escolher qualquer densidade? Impondo o desejo de parcimônia veremos que não é possível essa liberdade. Estamos interessados em não fazer nada. O aluno não deve se entusiasmar desnecessariamente. Se não chega informação, não deveria mudar nossa escolha de $p(x)$. Portanto a solução do problema variacional

$$\begin{aligned} \frac{\delta}{\delta p(x)}\left(S - \lambda \int p(x)dx\right) &= 0, \\ \dot{\Phi}\left(\frac{p(x)}{m(x)}, \frac{q(x)}{m(x)}\right) &= \lambda, \end{aligned} \tag{6.32}$$

deve levar à escolha $p(x) = q(x)$. Mas nem todo funcional do tipo 6.31 leva a esse resultado. Isto restringe mais a familia de funcionais. O problema 6.32 é uma equação diferencial ordinária

$$\begin{aligned} \dot{\Phi}(u,v) &= \lambda \\ \Phi(u,v) &= \lambda u + c(v) \end{aligned}$$

portanto, se seguissemos este caminho teriamos

$$\begin{aligned} S &= \int \left(\lambda \frac{p(x)}{m(x)} + c\left(\frac{q(x)}{m(x)}\right)\right) m(x)dx \\ &= \lambda \int p(x)dx + \text{const}, \end{aligned}$$

o que é uma trivialidade, pois a normalização diz que isto é uma constante e toda distribuição normalizada teria a mesma preferência. Portanto escolher a função Φ leva a um beco sem saida. Mas há outra saida para satisfazer a equação 6.32, basta escolher $m(x) = q(x)$, neste caso

$$\dot{\Phi}\left(\frac{p(x)}{q(x)}, \frac{q(x)}{q(x)}\right) = \lambda, \tag{6.33}$$

que é constante de forma trivial quando $p(x) = q(x)$. Novamente temos um avanço, restringindo ainda mais as estruturas possíveis:

$$S = \int \Phi\left(\frac{p(x)}{q(x)}\right) q(x)dx. \tag{6.34}$$

6.2.5 DE_5 : *Independência*

Este último desejo será usado para determinar a forma explícita de Φ na equação 6.34. Além da normalização, para o sistema composto pela garrafa de café, com graus de liberdade x_1 e a galáxia $X9$ com graus de liberdade x_2, temos os seguintes vínculos

$$\begin{aligned} \int p_1(x_1) f_1(x_1) dx_1 &= F_1 \\ \int p_2(x_2) f_2(x_2) dx_2 &= F_2. \end{aligned}$$

Tratando os dois sistemas de forma independente temos que $p_1(x_1)$ e $p_2(x_2)$ são dados pelas soluções de

$$\dot{\Phi}\left(\frac{p_1(x_1)}{q_1(x_1)}\right) = \lambda_{01} + \lambda_1 f_1(x_1) \text{ e } \dot{\Phi}\left(\frac{p(x_2)}{q_2(x_2)}\right) = \lambda_{02} + \lambda_2 f(x_2) \tag{6.35}$$

Tratando os sistemas juntos, o problema variacional é

$$\begin{aligned}
0 &= \frac{\delta}{\delta p(x_1,x_2)}\left\{S-\mu_0\left(\int p(x_1,x_2)dx-1\right)\right.\\
&- \mu_1\left(\int f_1(x_1)p(x_1,x_2)dx_1dx_2-F_1\right)\\
&- \left.\mu_2\left(\int f_2(x_2)p(x_1,x_2)dx_1dx_2-F_2\right)\right\},
\end{aligned}$$

$$\dot{\Phi}\left(\frac{p(x_1,x_2)}{q_1(x_1)q_2(x_2)}\right) = \mu_0+\mu_1 f_1(x_1)+\mu_2 f_2(x_2) \tag{6.36}$$

Impomos que a solução deve satisfazer $p(x_1,x_2)=p_1(x_1)p_2(x_2)$ e definimos $y=\frac{p_1(x_1)p_2(x_2)}{q_1(x_1)q_2(x_2)}$. Assim

$$\dot{\Phi}(y) = \mu_0+\mu_1 f_1(x_1)+\mu_2 f_2(x_2). \tag{6.37}$$

Derivamos duas vezes, primeiro com respeito a x_2

$$\ddot{\Phi}(y)\frac{\partial y}{\partial x_2} = (\mu_2 f_2(x_2))', \tag{6.38}$$

e depois com respeito a x_1

$$\dddot{\Phi}(y)\frac{\partial y}{\partial x_1}\frac{\partial y}{\partial x_2}+\ddot{\Phi}(y)\frac{\partial^2 y}{\partial x_1 x_2} = 0. \tag{6.39}$$

Isto parece complicado mas é bem simples. Note que, com a linha indicando a derivada com respeito a x_1 ou x_2

$$\begin{aligned}
\frac{\partial y}{\partial x_1} &= \left(\frac{p_1(x_1)}{q_1(x_1)}\right)'\frac{p_2(x_2)}{q_2(x_2)}\\
\frac{\partial y}{\partial x_2} &= \frac{p_1(x_1)}{q_1(x_1)}\left(\frac{p_2(x_2)}{q_2(x_2)}\right)'\\
\frac{\partial^2 y}{\partial x_1 x_2} &= \left(\frac{p_1(x_1)}{q_1(x_1)}\right)'\left(\frac{p_2(x_2)}{q_2(x_2)}\right)'.
\end{aligned} \tag{6.40}$$

Multiplique as duas primeiras e divida pela terceira equação, segue que $\Phi(y)$ satisfaz a equação diferencial ordinária:

$$y\dddot{\Phi}(y)+\ddot{\Phi}(y) = 0. \tag{6.41}$$

A solução geral contém três constantes arbitrárias

$$\Phi(y) = Ay\log y+By+C. \tag{6.42}$$

A familia de funcionais entropia é reduzida a

$$S[p||q]=A\int p(x)\log\frac{p(x)}{q(x)}dx+B\int p(x)dx+C\int q(x)dx.$$

A constante C pode ser tomada igual a zero, pois simplesmente adiciona uma constante sem mudar preferências. A constante B também pode ser tomada zero, pois a normalização de $p(x)$ será sempre um vínculo. Finalmente e demonstrando otimismo os fisicos tomam $A < 0$ negativo para que o processo variacional seja o de maximização. Podemos tomar $A > 0$ e minimizar o que é chamado de divergência de Kullback-Leibler. A escolha de $A = -1$ é simplesmente uma escolha de unidades da entropia. Como veremos isto levará a que a temperatura e a energia sejam medidos nas mesmas unidades. O resultado final é dado por

$$S[p||q] = -\int p(x) \log \frac{p(x)}{q(x)} dx \tag{6.43}$$

É facil mostrar que os resultados $p_1(x_1)p_2(x_2)$ usando a equação 6.35 e $p(x_1, x_2)$ da equação 6.36, coincidem.

Exercício Para variáveis discretas mostre que $H_{SJ} = -\sum_i p_i \log p_i$ difere por uma constante de $S[p||q] = -\sum_i p_i \log \frac{p_i}{q_i}$, se $q_i = 1/N$ é constante e portanto o valor das distribuições que as extremizam coincidem.

Isto permite concluir que a entropia de Shannon faz referência a uma ditribuição *a priori*, a distribuição uniforme. O fato de não aparentar de forma explícita não quer dizer que não esteja lá. É claro que o ponto onde a distribuição uniforme faz a sua entrada no raciocínio é quando a função $F(n)$ crescente é introduzida.

Exercício Suponha $\mathcal{F}$ uma família paramétrica de densidades, e dois membros $p_1 = p(x|\theta_1)$ e $p_2 = p(x|\theta_2)$ de $\mathcal{F}$. Suponha que os parâmetros θ estejam em algum subconjunto de $\mathbb{R}^K$. Mostre que em geral $S[p_1||p_2] \neq S[p_2||p_1]$, mas para valores pequenos da distância Euclideana, $|\theta_1 - \theta_2|$ há simetria de S ante $\theta_1 \leftrightarrow \theta_2$. Isso permite definir uma distância entre duas densidades e uma geometria Riemanniana no espaço dos θ, que agora passa a ser mais que um simples subconjunto de $\mathbb{R}^K$, mas o que é conhecido como uma variedade estatística. Considere $\theta_2 = \theta_1 + d\theta$ e expanda a entropia até segunda ordem em $d\theta$. Dessa forma é possível definir o tensor métrico $g_{\mu\nu}$. Encontre uma expressão para o tensor métrico que é conhecido pelo nome de Fisher-Rao. Encontre a distância entre duas gaussianas multivariada com a mesma covariância e médias diferentes.

6.3 *Apêndice: Multiplicadores de Lagrange*

Uma montanha é descrita pela altura $z = f(x, y)$, com x e y as coordenadas de um ponto de altura z e f um função suficientemente bem comportada, tem um máximo que pode ser obtido igulando as derivadas parciais de f a zero. Suponha que uma estrada cruza a superfície da montanha e as coordenadas da estrada são descritas por $\phi(x, y) = c$. Qual é a altura máxima de um carro que viaja pela estrada?

O problema de encontrar pontos extremos ou só estacionários de funções sujeitos a vínculos é muito vasto. Damos algumas idéias básicas sem preocupação com rigor, para lembrar o estudante de técnicas que deveriam ser vistas em Cálculo 2 ou curso equivalente.

Seja o problema

- P_{livre}: Queremos encontrar um ponto (x^*, y^*) dentro de uma certa região C no plano real onde uma função $f(x, y)$ tem localmente um valor estacionário.

Fácil, tome as derivadas parciais e resolva o sistema $\partial_x f = 0$, $\partial_y f = 0$.

Queremos, a seguir resolver um caso mais difícil.

- P_{vinc}: Suponha que não procuramos o resultado em qualquer lugar de C, mas especificamente queremos um ponto estacionário entre aqueles pontos que satisfazem $\phi(x, y) = c$, que supomos descreva uma curva no plano que está parcialmente contida em C e chamaremos γ.

A solução do parágrafo anterior dificilmente nos dá a resposta pois seria uma coincidência se (x^*, y^*) caisse encima dessa curva.

A solução a esta classe de problema foi proposta por Lagrange. Considere a classe de funções $F_\lambda(x, y)$ que dependem do chamado multiplicador de Lagrange λ:

$$F_\lambda(x, y) = f(x, y) + \lambda(\phi(x, y) - c) \tag{6.44}$$

Note que se o ponto de coordenadas x e y estiver na curva γ então F_λ e f tem o mesmo valor. Repetindo: F_λ e f tem o mesmo valor se o vínculo que " x e y estão sobre γ" for respeitado.

Consideremos o P_{livre} mas para a função $F_\lambda(x, y)$. O problema é novamente simples [6]. Resolvemos o sistema

[6] a não ser que não seja....

$$\frac{\partial F_\lambda}{\partial x} = 0 \tag{6.45}$$

$$\frac{\partial F_\lambda}{\partial y} = 0, \tag{6.46}$$

onde λ ainda não foi especificado. A resposta depende do valor escolhido para λ, isto é define uma curva ρ, parametrizada por λ : $(x^*(\lambda), y^*(\lambda))$ onde F_λ é extremo . Agora voltamos ao problema P_{vinc}. Da resposta à dupla pergunta "onde f é máximo?" e "onde o vínculo é satisfeito?" , quando as duas são respondidas simultaneamente, decorre a solução. Substituimos a primeira por "onde F é máximo?" (resposta: em ρ) junto com a afirmação "$f = F_\lambda$ sob a condição de estar em γ". Segue que queremos encontrar o cruzamento de γ e ρ. Basta escolhermos $\lambda = \lambda_*$ tal que $\phi\left(x^*(\lambda_*), y^*(\lambda_*)\right) = c$, o resultado é um extremo para f qunado se satisfaz o vínculo.

Agora procure um livro de cálculo e prencha os detalhes. Discuta também como lidar com casos em que o extremo está na borda de C. Há vínculos que são representados por desigualdades, os nomes de Kuhn e Tucker estão associados a esta extensão. Em muitos casos isto pode útil mas não no curso introdutório.