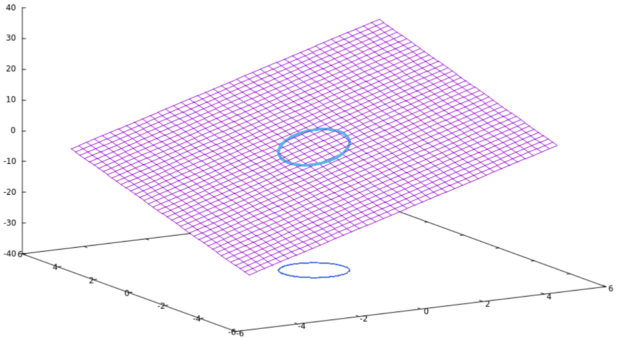
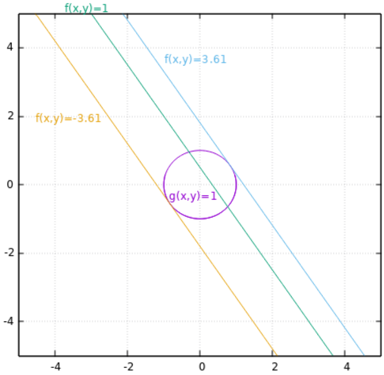
Zetknęliśmy się już niejednokrotnie z zagadnieniem poszukiwania ekstremum funkcji, a nawet, choć przelotnie, funkcjonału. Czasami jednak problem jest nieco bardziej skomplikowany - argumenty, dla których poszukujemy ekstremum, muszą spełniać jakieś dodatkowe warunki. Mówimy wtedy o optymalizacji z więzami, a ekstrema nazywamy ekstremami warunkowymi.

Weźmy przykładową funkcję $f(x, y) = 3x + 2y$ (płaszczyzna na rysunku). Znajdźmy jej minimum i maksimum... ale zaraz, jeśli poszukujemy ich w całej przestrzeni, ta funkcja przecież ich nie ma! Gdy jednak narzucimy warunek, że ekstremum ma spełniać dodatkową zależność: $g(x, y) = x^2 + y^2 = 1$, sytuacja może ulec zmianie - w końcu w obrębie jasnoniebieskiego okręgu znajdziemy wartość największą i najmniejszą.



Jak rozwiązać taki problem? Rzućmy okiem na tzw. wykres konturowy naszej funkcji, czyli punkty, dla których funkcja przyjmuje określoną wartość. Przykładowy wykres konturowy przedstawiony jest poniżej, a także (w wersji interaktywnej) w pliku „Lagrange.nb".



Zauważmy, że w minimum ($f(x, y) = -3.61$) i maksimum ($f(x, y) = 3.61$) wykresy konturowe f i g są styczne!

Dlaczego? Aby spełniać nałożony przez nas warunek, musimy ograniczyć poszukiwania do punktów spełniających równanie więzu, czyli do sytuacji, w których wykresy konturowe f i g mają jakieś wspólne punkty. Gdy po prostu przecinają się, wciąż możemy zwiększać lub zmniejszać funkcję f - czyli nie jesteśmy w ekstremum! Ekstremum znajdziemy tam, gdzie wykres konturowy f jedynie dotyka konturu g, czyli właśnie tam, gdzie są one ze sobą styczne. Co nam to daje? Gradient jest kierunkiem największego wzrostu funkcji, więc zawsze jest prostopadły do jej konturu (wzdłuż którego funkcja wcale się nie zmienia). W punktach, w których funkcje f i g są styczne, ich gradienty będą do siebie równoległe (zatem proporcjonalne). Pozwala to nam na bardziej formalne wyrażenie problemu:

$$\nabla f(x,y)|_{x_0,y_0} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

$$\nabla g(x,y)|_{x_0,y_0} = \begin{bmatrix} 2x_0 \\ 2y_0 \end{bmatrix}$$

$$\nabla f(x_0,y_0) = \lambda \nabla g(x_0,y_0)$$

λ to jakaś stała proporcjonalności pomiędzy gradientami f i g.

Czyli mamy:

$$\begin{aligned}
\nabla f(x_0, y_0) &= \lambda \nabla g(x_0, y_0) \\
3 &= \lambda 2 x_0 \\
2 &= \lambda 2 y_0 \\
x_0^2 + y_0^2 &= 1
\end{aligned}$$

Z tego układu 3 równań z 3 niewiadomymi w prosty sposób dostajemy: $x_0 = \frac{3}{2\lambda}, y_0 = \frac{1}{\lambda}$ oraz:

$$\begin{aligned}
\frac{9}{4\lambda^2} + \frac{1}{\lambda^2} &= 1 \\
\frac{9}{4} + 1 &= \lambda^2
\end{aligned}$$

Ostatecznie:

$$\begin{cases}
\lambda = \pm\frac{1}{2}\sqrt{13} \\
x_0 = \pm\frac{3}{\sqrt{13}} \\
y_0 = \pm\frac{2}{\sqrt{13}} \\
f(x_0, y_0) = \pm\sqrt{13} \approx \pm 3.61
\end{cases}$$

Lagrange wymyślił bardziej elegancki i uniwersalny sposób zapisu takiego rodzaju zagadnień, wprowadzając funkcję nazywaną dziś na jego cześć *lagranżjanem*[1]:

$$\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda(g(x, y) - c)$$

$g(x, y) = c$ to równanie więzu. Dlatego też wyraz stojący przy λ jest z w punktach stacjonarnych f z definicji równy 0[2]. Dzięki temu ekstrema warunkowe $\mathcal{L}$ i f będą wypadać w tych samych miejscach. Różniczkując lagranżjan po odpowiednich zmiennych, jesteśmy w stanie odzyskać wszystkie równania z poprzednich slajdów:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \lambda} &= -g(x, y) + c = 0 \\ g(x, y) &= c \\ \vec{0} = \nabla \mathcal{L} &= \nabla f - \lambda \nabla g \\ \nabla f &= \lambda \nabla g \end{aligned}$$

W ogólnym przypadku zapisuje się to podobnie, w postaci tzw. *równań Eulera* (j przebiega wszystkie zmienne niezależne, i - więzy):

$$\left[\left(\frac{\partial f}{\partial x_j} \right)_0 - \sum_i \lambda_i \left(\frac{\partial g}{\partial x_j} \right)_0 \right] = 0$$

[1] Nie mylić z lagranżjanem w mechanice klasycznej (różnicy energii kinetycznej i potencjalnej układu) - tamten lagranżjan to tak naprawdę szczególny przypadek tego tutaj (np. patrz „Mechanika klasyczna" Landaua i Lifszyca).

[2] Dodawanie zera lub mnożenie przez jeden to typowe sztuczki spotykane w matematyce, więc nie powinno nas to dziwić.

Spróbujmy wrócić do naszego zadania:

$$\mathcal{L}(x,y,\lambda) = 3x + 2y - \lambda(x^2 + y^2 - 1)$$
$$\frac{\partial \mathcal{L}}{\partial x} = 3 - \lambda 2x = 0$$
$$\frac{\partial \mathcal{L}}{\partial y} = 2 - \lambda 2y = 0$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = x^2 + y^2 - 1 = 0$$

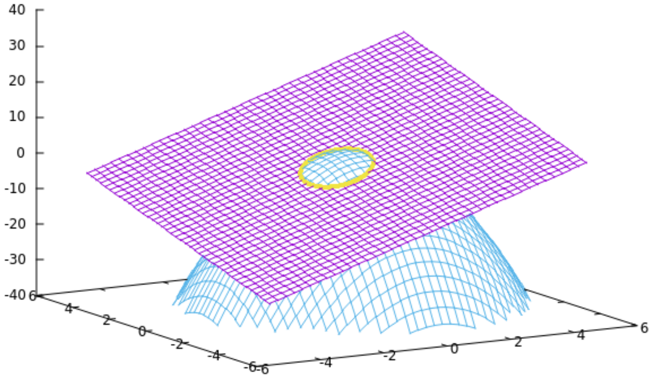
Czyli odzyskaliśmy wszystkie równania, jakie spełniać ma punkt stacjonarny. Możemy też równoważnie wyrazić ten problem nie w języku pochodnych, tylko różniczek:

$$d\mathcal{L}(x,y,\lambda) = 3dx + 2dy - 2\lambda x dx - 2\lambda y dy - (x^2 + y^2 - 1)d\lambda = 0$$
$$d\mathcal{L}(x,y,\lambda) = dx(3 - 2\lambda x) + dy(2 - 2\lambda y) - (x^2 + y^2 - 1)d\lambda = 0$$

Musi to być spełnione dla dowolnych $dx, dy, d\lambda$, więc nawiasy przyrównujemy do zera, odzyskując te same równania. Warto zauważyć, że często spotyka się nieco inne sformułowanie powyższych wzorów, gdzie jako zmienne traktuje się tylko x i y, a równanie więzu zapisuje się oddzielnie:

$$d\mathcal{L}(x,y) = 3dx + 2dy - 2\lambda x dx - 2\lambda y dy = 0$$
$$d\mathcal{L}(x,y) = dx(3 - 2\lambda x) + dy(2 - 2\lambda y) = 0$$
$$1 = x^2 + y^2$$

Na koniec zobaczmy jeszcze, jak wygląda wykres lagranżjanu nałożony na wykres funkcji f. Jak widać, dla spełnionego więzu funkcje te są sobie równe (i, jak wspomnieliśmy wcześniej, ich ekstrema warunkowe są sobie równe i wypadają w tych samych punktach).



Spróbujmy rozwiązać następujące zadanie: mamy pewien wektor: $\vec{v}$ i szukamy takiego wektora jednostkowego $\vec{u}^1$, który będzie nakładał się z nim maksymalnie (czyli szukamy maksimum iloczynu skalarnego).

$$\vec{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \vec{u}^1 = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \quad \vec{u}^1 \cdot \vec{u}^1 = x^2 + y^2 + z^2 = 1$$

Normalizacja wektora $\vec{u}^1$ jest tutaj więzem. Zapiszmy lagranżjan:

$$\begin{aligned} f(x,y,z) &= \vec{v} \cdot \vec{u}^1 = x + 2y + 3z \\ g(x,y,z) &= x^2 + y^2 + z^2 \\ c &= 1 \\ \mathcal{L}(x,y,z,\lambda) &= x + 2y + 3z - \lambda(x^2 + y^2 + z^2 - 1) \end{aligned}$$

Dostajemy, po zróżniczkowaniu:

$$x_0 = \frac{1}{2\lambda}, \quad y_0 = \frac{2}{2\lambda}, \quad z_0 = \frac{3}{2\lambda}$$

Bądź w zapisie wektorowym:

$$\vec{u}_0^1 = \begin{bmatrix} \frac{1}{2\lambda} \\ \frac{2}{2\lambda} \\ \frac{3}{2\lambda} \end{bmatrix} = \frac{1}{2\lambda} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \frac{1}{2\lambda} \vec{v}$$

Czyli, zgodnie z intuicją, wektory nakładają się maksymalnie wtedy, gdy są proporcjonalne (równoległe). Pozostało nam, z równania więzu, wyznaczyć nieoznaczony mnożnik λ:

$$1 = \frac{1}{4\lambda^2} + \frac{4}{4\lambda^2} + \frac{9}{4\lambda^2}$$
$$\lambda = \pm \frac{2}{\sqrt{7}}$$

Otrzymaliśmy również rozwiązanie ze znakiem minus - wówczas wektory są antyrównoległe i iloczyn skalarny przyjmuje najmniejszą możliwą wartość (najbardziej ujemną). Czyli znaleźliśmy dwa ekstrema - minimum i maksimum.

Jeszcze jeden przykład. Weźmy sytuację, którą rozważaliśmy na początku, tylko zmieńmy nieco płaszczyznę f:

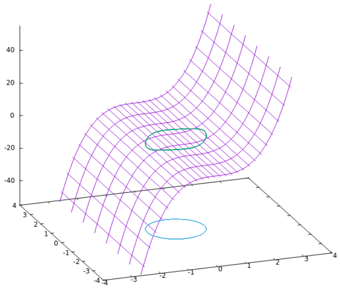
$$\begin{aligned} f(x,y) &= x^3 \\ g(x,y) &= x^2 + y^2 \\ c &= 1 \\ \mathcal{L}(x,y,\lambda) &= x^3 - \lambda(x^2 + y^2 - 1) \\ d\mathcal{L}(x,y,\lambda) &= dx(3x^2 - 2\lambda x) + dy(-2\lambda y) - (x^2 + y^2 - 1)d\lambda = 0 \end{aligned}$$

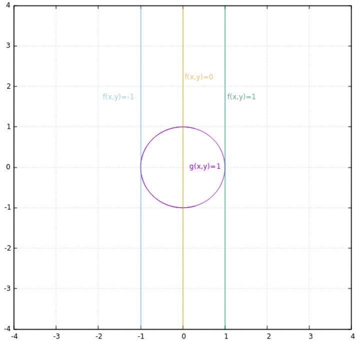
Stąd, z zerowania się nawiasów, dostajemy:

$$\begin{aligned} & 3x^2 - 2\lambda x = 0 \quad \rightarrow \quad x_0 = 0 \vee x_0 = \frac{2}{3}\lambda \\ & -2\lambda y = 0 \quad \rightarrow \quad y_0 = 0 \vee \lambda = 0 \\ & x^2 + y^2 - 1 = 0 \quad \rightarrow \quad \begin{cases} x_0 = 0 \\ y_0 = \pm 1 \\ \lambda = 0 \\ f(x_0, y_0) = 0 \end{cases} \vee \begin{cases} x_0 = \frac{2}{3}\lambda = \pm 1 \\ y_0 = 0 \\ \lambda = \pm\frac{3}{2} \\ f(x_0, y_0) = \pm 1 \end{cases} \end{aligned}$$

Czyli dostaliśmy 4 punkty - minimum, maksimum i dwa punkty przegięcia.

Jeszcze gwoli uzupełnienia - powyższa sytuacja graficznie:





W chemii kwantowej najbardziej interesować nas będzie zastosowanie metody nieoznaczonych mnożników Lagrange'a nie do funkcji, a do funkcjonałów. Analogicznie do pojęcia różniczki wprowadźmy pojęcie wariacji funkcjonału. Weźmy znany nam przykład funkcjonału - wartość oczekiwaną energii, która jest funkcjonałem zależnym od funkcji falowej ($E[\psi]$). Zastanówmy się, jak zmieni się ona gdy zmienimy funkcję falową:

$$\begin{aligned} E[\psi] &= \langle\psi|\hat{H}|\psi\rangle \\ E[\psi+\delta\psi] &= \langle\psi+\delta\psi|\hat{H}|\psi+\delta\psi\rangle = \langle\psi|\hat{H}|\psi\rangle + (\langle\delta\psi|\hat{H}|\psi\rangle + \langle\psi|\hat{H}|\delta\psi\rangle) + \langle\delta\psi|\hat{H}|\delta\psi\rangle \\ &= E[\psi] + \delta\,\langle\psi|\hat{H}|\psi\rangle + \langle\delta\psi|\hat{H}|\delta\psi\rangle = E[\psi] + \delta E + \ldots \end{aligned}$$

Wyrażenie w nawiasie okrągłym (liniowe w $\delta\psi$) zapisaliśmy jako δE i nazywamy je (pierwszą) wariacją funkcjonału $E[\psi]$. Ostatni wyraz jest kwadratowy w $\delta\psi$ i nie będziemy się nim zajmować. Warto zauważyć, że δ działa analogicznie do operatora różniczkowania (w ten sposób zwinęliśmy wyrażenie w nawiasie okrągłym). Poszukiwanie minimum (a w zasadzie ekstremum) funkcjonału (patrz: metoda wariacyjna) możemy rozumieć jako znalezienie takiego ψ, dla którego pierwsza wariacja E się zeruje (czyli $\delta E = 0$). Dla funkcjonału również możemy zapisać lagranżjan:

$$\mathcal{L}[\psi] = f[\psi] - \lambda(g[\psi] - c)$$
$$\delta\mathcal{L} = \delta f - \lambda\delta g = 0$$

Sprawdźmy, jak wygląda wyrażone w ten sposób wyprowadzenie metody wariacyjnej Ritza. Jako więz traktujemy tu[3] normalizację $|\psi\rangle$.

$$
\begin{aligned}
|\psi\rangle &= \sum_i c_i \, |\phi_i\rangle \\
f &= \langle\psi|\hat{H}|\psi\rangle \,, \qquad g = \langle\psi|\psi\rangle \,, \qquad c = 1 \\
\mathcal{L} &= \langle\psi|\hat{H}|\psi\rangle - \lambda(\langle\psi|\psi\rangle - 1) \\
&= \sum_{ij} c_i^* c_j \, \langle\phi_i|\hat{H}|\phi_j\rangle - \lambda(\sum_{ij} c_i^* c_j \, \langle\phi_i|\phi_j\rangle - 1) \\
\delta\mathcal{L} &= \delta \, \langle\psi|\hat{H}|\psi\rangle - \lambda\delta \, \langle\psi|\psi\rangle \\
&= \sum_{ij} \delta c_i^* c_j \, \langle\phi_i|\hat{H}|\phi_j\rangle - \lambda \sum_{ij} \delta c_i^* c_j \, \langle\phi_i|\phi_j\rangle \\
&+ \sum_{ij} c_i^* \delta c_j \, \langle\phi_i|\hat{H}|\phi_j\rangle - \lambda \sum_{ij} c_i^* \delta c_j \, \langle\phi_i|\phi_j\rangle \\
&= \sum_{ij} \delta c_i^* c_j \, [H_{ij} - \lambda S_{ij}] \\
&+ \sum_{ij} \delta c_j c_i^* \, [H_{ij} - \lambda S_{ij}] = 0
\end{aligned}
$$

[3] Jak w przypadku zadania z wektorami i jak później, przy wyprowadzaniu metody Hartree-Focka.

Wskaźniki i i j przebiegają dokładnie ten sam zakres. Zamieńmy je więc w drugiej części wzoru:

$$\delta\mathcal{L} = \sum_{ij} \delta c_i^* c_j \left[H_{ij} - \lambda S_{ij}\right] + \sum_{ji} \delta c_i c_j^* \left[H_{ji} - \lambda S_{ji}\right] = 0$$

Macierze $\boldsymbol{H}$ i $\boldsymbol{S}$ są hermitowskie ($M_{ij} = M_{ji}^*$), ponadto λ (jako stała proporcjonalności pomiędzy rzeczywistymi $\delta \langle\psi|\hat{H}|\psi\rangle$ i $\langle\psi|\psi\rangle$) jest rzeczywista. Dlatego też drugą część wzoru możemy zapisać w nieco inny sposób (dla czytelności oddzielnie zapisaliśmy też sumy):

$$\begin{aligned}\delta\mathcal{L} &= \sum_i \delta c_i^* \sum_j c_j \left[H_{ij} - \lambda S_{ij}\right] + \sum_i \delta c_i \sum_j c_j^* \left[H_{ij}^* - \lambda S_{ij}^*\right] \\ &= \sum_i \delta c_i^* \sum_j \left\{c_j \left[H_{ij} - \lambda S_{ij}\right]\right\} + \sum_i \delta c_i \sum_j \left\{c_j \left[H_{ij} - \lambda S_{ij}\right]\right\}^* = 0\end{aligned}$$

...dla dowolnych δc_i i δc_i^*.

Skoro otrzymane równanie ma być spełnione dla dowolnych δc_i i δc_i^*, zerować się muszą nawiasy klamrowe, czyli:

$$c_j \left[H_{ij} - \lambda S_{ij} \right] = 0$$
$$(\boldsymbol{H} - \lambda \boldsymbol{S})\boldsymbol{c} = \boldsymbol{0}$$

Oczywiście jeśli powyższe jest równe 0, to także jego sprzężenie zespolone jest zerowe — zatem z zerowania się drugiego nawiasu klamrowego dostajemy ten sam wzór! W rezultacie w podobnych sytuacjach mamy prawo (teraz już uzasadnione) wariować tylko po współczynnikach δc_i albo tylko po δc_i^*. Dostaliśmy zatem znane nam już wyrażenie, kluczowe dla metody wariacyjnej Ritza, zaś λ ma fizyczny sens przybliżenia do energii oczekiwanej (wcześniej oznaczaliśmy to jako ε). Ogólnie patrząc, dość często okazuje się, że nieoznaczone mnożniki mają proste interpretacje fizyczne.