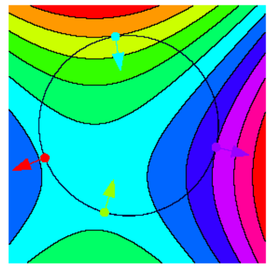
**CONSTRAINED EXTREMA** **Math21a, O. Knill**

HOMEWORK: 13.9: 4,12,28,24,34

CONSTRAINED EXTREMA. Given a function $f(x,y)$ of two variables and a level curve $g(x,y) = c$. Find the extrema of $f$ on the curve. You see that at places, where the gradient of $f$ is not parallel to the gradient of $g$, the function $f$ changes when we change position on the curve $g = c$. Therefore we must have a solution of three equations

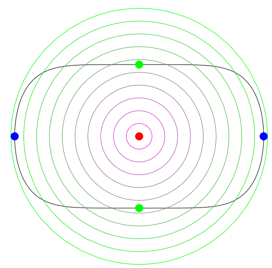
$$\boxed{\nabla f(x,y) = \lambda \nabla g(x,y), g(x,y) = c}$$

to the three unknowns $(x,y,\lambda)$. (Additionally: check points with $\nabla g(x,y) = (0,0)$). The constant $\lambda$ is called the **Lagrange multiplier**. The equations obtained from the gradients are called **Lagrange equations**.



EXAMPLE. To find the shortest distance from the origin to the curve $x^6 + 3y^2 = 1$, we extremize $f(x,y) = x^2 + y^2$ under the constraint $g(x,y) = x^6 + 3y^2 - 1 = 0$.

SOLUTION: $\nabla f = (2x, 2y), \nabla g = (6x^5, 6y)$. $\nabla f = \lambda \nabla g$ gives the system $2x = \lambda 6x^5, 2y = \lambda 6y, x^6 + 3y^2 - 1 = 0$. We get $\lambda = 1/3, x = x^5$, so that either $x = 0$ or $1$ or $-1$. From the constraint equation, we obtain $y = \sqrt{(1-x^6)/3}$. So, we have the solutions $(0, \pm\sqrt{1/3})$ and $(1,0), (-1,0)$. To see which is the minimum, just evaluate $f$ on each of the points. We see that $(0, \pm\sqrt{1/3})$ are the minima.



HIGHER DIMENSIONS. The above constrained extrema problem works also in more dimensions. For example, if $f(x,y,z)$ is a function of three variables and $g(x,y,z) = c$ is a surface, we solve the system of 4 equations

$$\boxed{\nabla f(x,y,z) = \lambda \nabla g(x,y,z), g(x,y,z) = c}$$

to the 4 unknowns $(x,y,z,\lambda)$. If we want to extremize $f(x,y,z)$ under two constraints $g(x,y,z) = c$ and $h(x,y,z) = d$, we have a system of 5 equations for 5 unknowns:

$$\boxed{\nabla f(x,y,z) = \lambda \nabla g(x,y,z) + \mu \nabla h(x,y,z), g(x,y,z) = c, h(x,y,z) = d}$$

EXAMPLE. Extrema of $f(x,y,z) = z$ on the sphere $g(x,y,z) = x^2 + y^2 + z^2 = 1$ are obtained by calculating the gradients $\nabla f(x,y,z) = (0,0,1)$, $\nabla g(x,y,z) = (2x, 2y, 2z)$ and solving $(0,0,1) = \nabla f = \lambda \nabla g = (2\lambda x, 2\lambda y, 2\lambda z), x^2 + y^2 + z^2 = 1$. $\lambda = 0$ is excluded by the third equation $1 = 2\lambda z$ so that the first two equations $2\lambda x = 0, 2\lambda y = 0$ give $x = 0, y = 0$. The 4'th equation gives $z = 1$ or $z = -1$. The extrema are the north pole $(0,0,1)$ (maximum) and the south pole $(0,0,-1)$ (minimum).

THE PRINCIPLE OF MAXIMAL ENTROPY. Consider a dice showing $i$ with probability $p_i$, where $i = 1, ..., 6$. The **entropy** of the probability distribution is defined as $S(\vec{p}) = -\sum_{i=1}^{6} p_i \log(p_i)$. Find the distribution $p$ which maximizes entropy under the constrained $g(\vec{p}) = \sum_{i=1}^{6} p_i = 1$.

SOLUTION: $\nabla f = (-1 - \log(p_1), \dots, -1 - \log(p_n))$, $\nabla g = (1, \dots, 1)$. The Lagrange equations are $-1 - \log(p_i) = \lambda, p_1 + ... + p_6 = 1$, from which we get $p_i = e^{-(\lambda+1)}$. The last equation $1 = \sum_i \exp(-(\lambda+1)) = 6\exp(-(\lambda+1))$ fixes $\lambda = -\log(1/6) - 1$ so that $p_i = 1/6$. The distribution, where each event has the same probability is the distribution of maximal entropy.

REMARK. Maximal entropy means **least information content**. A dice which is fixed (asymmetric weight distribution for example) allows a cheating gambler to gain profit. Cheating through asymmetric weight distributions can be avoided by making the dices transparent.

THE PRINCIPLE OF MINIMAL FREE ENERGY. Assume that the probability that a system is in a state $i$ is $p_i$ and that the energy of the state $i$ is $E_i$. Nature tries to minimize the **free energy** $f(p_1, ..., p_n) = -\sum_i (p_i \log(p_i) - E_i p_i)$ when the energies $E_i$ are fixed. The **free energy** is the difference of the **entropy** $S(p) = -\sum_i p_i \log(p_i)$ and the **energy** $E(p) = \sum_i E_i p_i$. The probability distribution $p_i$ satisfying $\sum_i p_i = 1$ minimizing the free energy is called the **Gibbs distribution**.

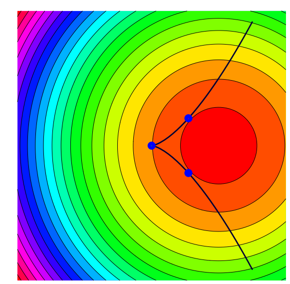
SOLUTION: $\nabla f = (-1 - \log(p_1) - E_1, \dots, -1 - \log(p_n) - E_n)$, $\nabla g = (1, \dots, 1)$. The Lagrange equation are $\log(p_i) = -1 - \lambda - E_i$, or $p_i = \exp(-E_i)C$, where $C = \exp(-1-\lambda)$. The additional equation $p_1 + \dots + p_n = 1$ gives $C(\sum_i \exp(-E_i)) = 1$ so that $C = 1/(\sum_i e^{-E_i})$. The Gibbs solution is $p_i = \exp(-E_i)/\sum_i \exp(-E_i)$. For example, if $E_i = E$, then the Gibbs distribution is the uniform distribution $p_i = 1/n$.

MOST ECONOMIC ALUMINUM CAN. You manufacture cylindrical soda cans of height $h$ and radius $r$. You want for a fixed volume $V(r,h) = h\pi r^2 = 1$ a minimal surface area $A(r,h) = 2\pi rh + 2\pi r^2$. With $x = h\pi, y = r$, you need to optimize $f(x,y) = 2xy + 2\pi y^2$ under the constrained $g(x,y) = xy^2 = 1$. Calculate $\nabla f(x,y) = (2y, 2x + 4\pi y), \nabla g(x,y) = (y^2, 2xy)$. The task is to solve $2y = \lambda y^2, 2x + 4\pi y = \lambda 2xy, xy^2 = 1$. The first equation gives $y\lambda = 2$. Putting that in the second one gives $2x + 4\pi y = 4x$ or $2\pi y = x$. The third equation finally reveals $2\pi y^3 = 1$ or $y = 1/(2\pi)^{1/3}, x = 2\pi(2\pi)^{1/3}$. This means $h = 0.54.., r = 2h = 1.08$.
REMARK. Other factors can influence the shape. For example, the can has to withstand a pressure up to 100 psi.

WHERE DO CONSTRAINED EXTREMAL PROBLEMS OCCUR. (informal)

- Constraints occur at boundaries. If we want to maximize a function over a region, we have also to look at the extrema at the boundary, where the gradient not necessarily vanishes.
- In physics, we often have conserved quantities (like for example energy, or momentum, or angular momentum), constraints occur naturally.
- In economical contexts, one often wants to optimize things under constraints.
- In mechanics constraints occur naturally (i.e. for robot arms).
- Constrained optimization is important in statistical mechanics, where equilibria have to be obtained under constraints (i.e. constant energy).
- Probability distributions are solutions of constrained problems. For example, the Gaussian distribution (normal distribution) is the distribution on the line with maximal entropy.
- Eigenvalue problems in linear algebra can be interpreted as constrained problems. Maximizing $u \cdot Lu$ under the condition $|u|^2 = 1$ gives the equation $Lu = \lambda u$ and $u$ has to be an eigenvector. The Lagrange multiplier is an eigenvalue.

DID WE GET ALL EXTREMA? The Lagrange equations $\nabla f(x,y) = \lambda \nabla g(x,y), g(x,y) = c$ do not give all the extrema. We can also have $\nabla g(x,y) = (0,0)$. The parallelity condition also could have been written as $\lambda \nabla f(x,y) = \nabla g(x,y)$, $f(x,y) = x^2 + (y-1)^2, g(x,y) = x^2 - y^3$. The function $f$ has a local maximum 1 at $(0,0)$ under the constraint $g(x,y) = 0$ but the Lagrange equations do not find it. The problem is that the gradient of $g$ vanishes. $\nabla g$ is technically parallel to $\nabla f$ but there is no $\lambda$ such that $\nabla f = \lambda \nabla g$ at this point. The reason for this "mistake" (which is present in virtually all calculus text books), is that parallelity of the two gradient is not equivalent to $\nabla f = \lambda \nabla g$ but can also mean $\lambda \nabla f = \nabla g$ with $\lambda = 0$.



CAN WE AVOID LAGRANGE? We could extremize $f(x,y)$ under the constraint $g(x,y) = 0$ by finding $y = y(x)$ from the later and extremizing the 1D problem $f(x, y(x))$.

EXAMPLE 1. To extremize $f(x,y) = x^2 + y^2$ with constraint $g(x,y) = x^4 + 3y^2 - 1 = 0$, solve $y^2 = (1 - x^4)/3$ and minimize $h(x) = f(x, y(x)) = x^2 + (1 - x^4)/3$. $h'(x) = 0$ gives $x = 0$. The find the maximum $(\pm 1, 0)$, we had to maximize $h(x)$ on $[-1, 1]$, which occurs at $\pm 1$.

**Sometimes, the Lagrange method can be avoided**.

EXAMPLE 2. Extremize $f(x,y) = x^2 + y^2$ under the constraint $g(x,y) = p(x) + p(y) = 1$, where $p$ is a complicated function in $x$ which satisfies $p(0) = 0, p'(1) = 2$ The Lagrange equations $2x = \lambda p'(x), 2y = \lambda p'(y), p(x) + p(y) = 1$ can be solved: with $x = 0, y = 1, \lambda = 1$, however, we can not solve $g(x,y) = 1$ for $y$. Substitution fails:
**In general, the Lagrange method is more powerful.**