SECOND ORDER OPTIMALITY CONDITIONS

~~Recall that~~ Definition: self-adjoint operator (symmetric matrix) $A \in Mat(n \times n, \mathbb{R})$ is <u>positive semidefinite</u> if $v^T A v = (Av, v) = \sum a_{k\ell} v_k v_\ell \geq 0$ for all $v \in \mathbb{R}^n$.

<u>positive definite</u> if $(Av, v) > 0$ for any $v \neq 0$.

<u>Thm</u> If $f : \mathcal{U} \to \mathbb{R}$ (open) is twice Gâteaux differentiable, and $x \in \mathcal{U}$ is a local minimizer, then $Hf = \left(\frac{\partial^2 f}{\partial x_k \partial x_\ell}\right)_{k,\ell = 1..n}$ is positive-semidefinite (and, of course, $\nabla f = 0$).

<u>Proof:</u> Taylor formula along lines $(x + tv)_{t \in \mathbb{R}}$ $\square$

NB: of course, this is just a necessary, not sufficient condition of minimality:

$f(x) = x^3$; $\nabla f = 0$, $Hf = 0$ (pos. semidef), but not local minimizer.

Sufficient condition:

<u>Thm</u> $f : \mathcal{U} \to \mathbb{R}$; two times Fréchet differentiable, $\nabla f(x) = 0$, $Hf(x)$ is positive-definite. Then x is a local minimizer

<u>Proof:</u> Taylor formula! + the fact that if H is pos. def., then for some $c > 0$, $(Hv, v) \geq \alpha |v|^2$.

Proof of : consider the set $\{|v| = 1\}$ — it is closed & bounded in $\mathbb{R}^n \Rightarrow$ compact. Hence (Hv, v) attains its minimum $c > 0$.

But $(Hv, v) = \left(H\frac{v}{|v|}, \frac{v}{|v|}\right) \cdot |v|^2 \geq c|v|^2$ $\square$ $\square$

Another easy but useful result: If f is positive <u>semi</u> definite on open <u>convex</u> $\mathcal{U}$, and $\nabla f(x) = 0$, then x is a global minimizer on $\mathcal{U}$.

Indeed, $f(y) = f(x) + \underbrace{(\nabla f(x), y - x)}_{=0} + \underbrace{\left(\frac{Hf(z)}{2} y - x, y - x\right)}_{\geq 0}$ for some $z \in (x, y)$

$\square$

NB: If Hf is <u>indefinite</u>, then a critical pt x is a saddle.

Definition Quadratic form given by a matrix A is called nondegenerate if $\det A \neq 0$.

A critical pt a whose Hf is nondegenerate is called Morse.

Examples $f = 3xy - x^2y - xy^2$

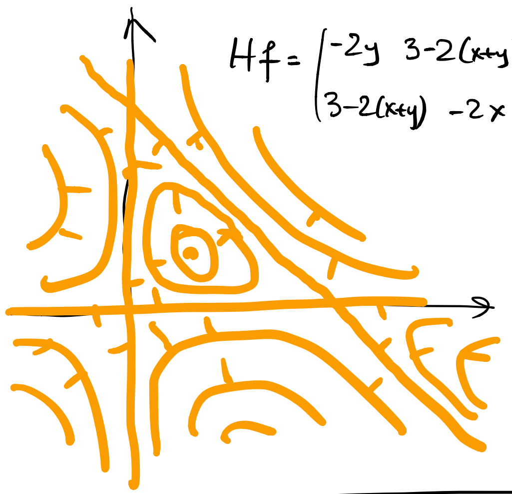
$$\nabla f = (3y - 2xy - y^2,\ 3x - x^2 - 2xy)$$

Critical points:

$(0,0)$ $(1,1)$ $(0,3)$ $(3,0)$

saddle

local maximum



$$Hf = \begin{pmatrix} -2y & 3-2(x+y) \\ 3-2(x+y) & -2x \end{pmatrix}$$

$$H = \begin{pmatrix} -2 & -1 \\ -1 & -2 \end{pmatrix} \qquad (Hv, v) =$$

$$= -2v_1^2 - 2v_1v_2 - 2v_2^2 =$$

$$= -v_1^2 - (v_1+v_2)^2 - v_2^2 \ll 0$$

Systematic understanding of the structure of functions near critical pts requires analysis of quadratic forms

$$q(v) = (Hv, v) = \sum a_{k\ell} v_k v_\ell$$

Easiest quadratic forms: diagonal ones; where

$$H = \begin{pmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \ \lambda_n \end{pmatrix}, \text{ so that } q(v) = \sum \lambda_k v_k^2$$

NB: $\{\lambda_1 \leq \ldots \leq \lambda_n\} = \sigma(H)$ spectrum of H.

Clearly, if all $\lambda_k \geq 0$ (>0), $k = 1, \ldots, n$ then q is positive semidefinite (pos. definite).

if some $\lambda_k < 0 < \lambda_\ell$, then q is a saddle.

All ≤ 0 (< 0) — negative semindefinite (def.)

Remarkably, any symmetric matrix can be turned into diagonal one by a change of coordinates...

[e.g. $2x^2-2xy+2y^2=3\left(\frac{x+y}{\sqrt{2}}\right)^2+\left(\frac{x-y}{\sqrt{2}}\right)^2$ — remark that vectors $(1/\sqrt{2},1/\sqrt{2})$, $(1/\sqrt{2},-1/\sqrt{2})$ form an orthonormal basis]

Namely, if $q=\sum a_{kl}x_k x_l$, then there is a unitary matrix U [unitary matrices are characterized by $U\cdot U^T=E$, or, equivalently by the property that their rows — or columns — form an orthonormal basis] such that for coordinates defined by $x=Uy$, the resulting form $\tilde{q}(y)=q(Uy)=(AUy,Uy)=U^TAU=\Sigma$ is diagonal.

The eigenvalues of a matrix do not change under similarities $X\rightsquigarrow U^{-1}XU$, so the spectra of A and $U^TAU=U^{-1}AU$ are the same. $\sigma=\{\lambda_1,\dots,\lambda_n\}\leftsquigarrow\begin{pmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_n\end{pmatrix}$

[Recall, by the way, that spectra of symmetric matrices are real]

Last useful fact: given 2 symmetric matrices, A & B, such that A is positive definite, there is a coordinate transformation $x=Xy$, such that both A & B become diagonal, $\Sigma_1=X^TAX$, $\Sigma_2=X^TBX$.

Proof: Declare (Ax,x) new Euclidean norm.

In more details: One can find U orthogonal: $U^TAU=\begin{pmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_n\end{pmatrix}$, $\lambda_k>0$ $k=1..n$.

Then for $C:=U\begin{pmatrix}\sqrt{\lambda_1} & & 0\\ & \ddots & \\ 0 & & \sqrt{\lambda_n}\end{pmatrix}U^T$ one has: $C^{-1}AC=E$. Then $C^{-1}BC^{-1}=:\overline{B}$ is symmetric (as $C^T=C$) and can be represented as $V\Sigma_2V^T$ for some orthogonal V. Set $X:=C^{-1}V$. Then $X^TBX=V^TC^{-1}BC^{-1}V=\Sigma_2$, $X^TAX=V^TC^{-1}AC^{-1}V=E$ ...

$\square$

SYLVESTER'S CRITERION

How to determine, whether a symmetric matrix A generates a positive-definite quadratic form $q(v) = (Av, v)$?

$$A = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & & a_{2n} \\ \vdots & & \ddots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

Determinants of $\square$ are called (principal) minors

Thm: (Av, v) is positive-definite iff all nested principal minors are > 0.

Proof $\Rightarrow$ is immediate: if $(Av, v) > 0$ for all v, it is also on subspace $(v_1 \ldots v_k, 0, 0, \ldots 0)$. So quadratic form on $\mathbb{R}^k$ given by $A_k = \begin{pmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & & \\ a_{k1} & & a_{kk} \end{pmatrix}$ is positive definite, and so all $\lambda_\ell \in \sigma(A_k)$ are > 0, so $\det A_k > 0$ $\square$

$\Leftarrow$. Assume it's true in dim n, prove for $n+1$. (Base, $n = 1$, is trivial)

$A = \begin{vmatrix} B & b \\ b^T & c \end{vmatrix}$ $B \in Mat(n \times n)$, $b \in Mat(n \times 1)$ ...

If A is not pos. def, there is $\lambda_{n+1} < 0$. But as (by assumption) $\det A > 0$, there is $\lambda_n < 0$. Let $Au = \lambda_n u$, $AV = \lambda_{n+1} V$.

Now, some linear combination $W = \mu u + \nu V$ has last coordinate $= 0$ (but $w \neq 0$). Then $(Aw, w) = (\mu \lambda_n u + \nu \lambda_{n+1} V, \mu u + \nu V) =$

$$= \lambda_n \mu^2 (u, u) + \lambda_{n+1} \nu^2 (v, v) < 0.$$

But that means that B is not positive definite, contradiction $\square$

Example $\begin{vmatrix} 2 & 1 & & & & \\ 1 & 2 & 1 & & 0 & \\ & 1 & 2 & 1 & & \\ & & 1 & 2 & \ddots & \\ & 0 & & \ddots & \ddots & 1 \\ & & & & 1 & 2 \end{vmatrix}$ is positive definite.