LECTURE 2 - 9/13/22

Parabolas are everywhere in nature and otherwise, which are examples of quadratic functions.

Examples: Linear regressions, matrix factorization, image processing.
$\min \frac{1}{2}\sum (y_i-\hat{y}_i)^2$

Today we discuss quadratic functions very broadly.

- Quadratic functions in n dimensions
- Gradient, extreme points
- Concave/convex/Saddle geometry
- Spectral Decomposition Theorem
- Linear approximation of any continuous differentiable functions. } Taylor's Thm.
- Hessian matrix and quadratic approximation of any continuous differentiable function.

Quadratic Functions

one dimension: $f:\mathbb{R}\to\mathbb{R}$, $f(x)=ax^2+bx+c$, $a,b,c\in\mathbb{R}$

$a>0 \to$ Convex/opens up, $a<0 \to$ concave/opens down

To min or max solve $f'(x)=0$. vertex $\Rightarrow \left(-\frac{b}{2a}, f\left(-\frac{b}{2a}\right)\right)$.

two dimensions: $f:\mathbb{R}^2\to\mathbb{R}$ $f(x_1,x_2)=a_{11}x_1^2+a_{22}x_2^2+a_{12}x_1x_2 + b_1x_1+b_2x_2+c$

Convex $\to$ min exists Concave $\to$ max exists
saddle shape $\to$ no min or max.

we introduce matrix notation.

$x=\begin{pmatrix}x_1\\x_2\end{pmatrix}$, $A=\begin{pmatrix}a_{11} & a_{12}\\ 0 & a_{22}\end{pmatrix}$, $b=\begin{pmatrix}b_1\\b_2\end{pmatrix}$, $c\in\mathbb{R}$ $\Rightarrow f(x)=x^TAx+b^Tx+c$.

Another way to represent A is if $\frac{a_{21}+a_{12}}{2}$ is the coef of the cross product term x_1x_2. The matrix form given above is not unique.

Similarly to the 2-dim case, we have 3-dim quadratic represented as $f(x) = \vec{x}^T A \vec{x} + \vec{b}^T \vec{x} + c$ in 3-D rather than 2.

Ex: $f(x) = 10x_1^2 + 3x_2^2 + 5x_1x_2 - x_2x_3 + 10x_1$

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad A = \begin{pmatrix} 10 & \boxed{4} & 0 \\ \boxed{1} & 3 & 2 \\ 0 & -3 & 0 \end{pmatrix} \quad b = \begin{pmatrix} 10 \\ 0 \\ 0 \end{pmatrix}.$$

Ex: $f(x) = 5x_1^2 - 6x_3^2 + 5x_1x_2 - 2x_2x_3 - x_1 - x_2 - x_3 + 100$

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad A = \begin{pmatrix} 5 & 2.5 & 0 \\ 2.5 & 0 & -1 \\ 0 & -1 & -6 \end{pmatrix} \quad b = \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix} \quad c = 100$$

This will be an exam problem.

Generalization: n-dimensional quadratic $f: \mathbb{R}^n \to \mathbb{R}$

Still use $f(x) = \vec{x}^T A \vec{x} + \vec{b}^T \vec{x} + c$.

we prefer, without loss of generality, and assume A symmetric. i.e. $A = A^T$.

Extreme points: points where function attains min/max/saddle. Where derivative/gradient/partials are all zero (depend on dim). Horizontal tangent (hyper-) plane.

To determine extreme points, solve $\nabla f(x) = 0$, $\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$

Finding gradient of n-dimensional Quadratic

$\nabla f(x)$, $f(x) = \vec{x}^T A \vec{x} + \vec{b}^T \vec{x} + c$, $A = A^T$

Easier to shift perspective to inner products:

$f(x) = \langle \vec{x}, A\vec{x} \rangle + \langle \vec{b}, \vec{x} \rangle + c$. Then:

$$\nabla f(x) = \begin{pmatrix} \partial f/\partial x_1 \\ \vdots \\ \partial f/\partial x_n \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial x_1}(\langle \vec{x}, A\vec{x} \rangle + \langle \vec{b}, \vec{x} \rangle + c) \\ \vdots \\ \frac{\partial}{\partial x_n}(\langle \vec{x}, A\vec{x} \rangle + \langle \vec{b}, \vec{x} \rangle + c) \end{pmatrix} = \underbrace{\begin{pmatrix} \frac{\partial}{\partial x_1}(\langle \vec{x}, A\vec{x} \rangle) \\ \vdots \end{pmatrix} + \begin{pmatrix} \frac{d}{dx_1}\langle \vec{b}, \vec{x} \rangle \\ \vdots \end{pmatrix} + \begin{pmatrix} d/dx_1\, c \\ \vdots \end{pmatrix}}_{\text{sum of derivs}}$$

$\frac{\partial}{\partial x_i}(c) = 0 \Rightarrow \nabla(c) = 0$

$$\frac{\partial}{\partial x_i}(\langle \vec{b}, \vec{x} \rangle) = \frac{\partial}{\partial x_i}(b_1 x_1 + \cdots + b_i x_i + \cdots + b_n x_n) = b_i$$

$$\Rightarrow \nabla(\langle \vec{b}, \vec{x} \rangle) = \begin{pmatrix} \frac{\partial}{\partial x_1}(\langle \vec{b}, \vec{x} \rangle) \\ \vdots \\ \frac{\partial}{\partial x_n}(\langle \vec{b}, \vec{x} \rangle) \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \vec{b}$$

Similarly:

$\nabla(\langle \vec{x}, A\vec{x} \rangle) = (A + A^T)\vec{x}$ (dot product expansion

If $A = A^T$, then $\nabla(\langle \vec{x}, A\vec{x} \rangle) = 2A\vec{x}$

so $\nabla f(x) = 2A\vec{x} + \vec{b}$. Then solve $\nabla f(\vec{x}) = 0$ for extreme point.

$\nabla f(\vec{x}) = 0 \rightarrow \vec{x} = -\frac{1}{2}A^{-1}\vec{b}$ (If A is invertible).

If A is not $A^T = A$, then $\vec{x} = -(A + A^T)^{-1}\vec{b}$

we can do optimization using linear algebra.

Ex: Find the extreme point: $f(x) = \alpha\|x\|^2 + \|Ax - b\|^2$, $\alpha > 0$.

$= \alpha\langle x, x \rangle + \langle Ax - b, Ax - b \rangle$

Note: we assume L2 Norm with no notation o/w.

$f(x) = \alpha\langle x, x \rangle + \langle Ax - b, Ax - b \rangle = \alpha\langle x, x \rangle + \langle Ax, Ax \rangle - 2\langle Ax, b \rangle + \langle b, b \rangle$.

$= \alpha\langle x, x \rangle + \langle x, A^T A x \rangle - 2\langle b, Ax \rangle + \langle b, b \rangle$.

$$\Rightarrow f(x) = x^T \underbrace{(\alpha I + A^T A)}_{Q} x - \underbrace{2b^T A}_{d^T} x + \underbrace{\|b\|^2}_{c}$$

$= x^T Q x - d^T x + c$ ← Quadratic form.

Q is symmetric, so $\nabla f = 2Qx + d = 0 \Rightarrow x = -\frac{1}{2}Q^{-1}d$

How do we know if convex or concave or saddle? →

General Rule: $f(x) = x^T A x + b^T x + c$

If $x^T A x > 0 \ \forall x \neq 0$ A is called <u>positive definite</u>, denoted by $A \succ 0$.
All eigenvalues ~~are~~ of A are positive. $\Rightarrow$ f convex, has min point.

If $x^T A x < 0 \ \forall x \neq 0$ A is called negative definite, denoted $A \prec 0$.
All eigenvalues of A are negative. $\Rightarrow$ f concave, has max point.

If neither, A is not positive/negative definite, so f has a saddle point.

Can we change the ~~basis~~ function so that the cross terms are eliminated? $\rightarrow$ Yes, if A is symmetric this can be done by changing the basis.

<u>Spectral Decomposition Thm</u>

<u>THM</u> If A is symmetric $n \times n$, there exist n eigenvectors $v_1, \dots, v_n \in \mathbb{R}^n$ with real eigenvalues $\lambda_1, \dots, \lambda_n$ such that

- $\langle v_i, v_j \rangle = 0$ if $i \neq j$
- $\|v_i\| = 1$.

$\Big\}$ That is, $v_1, \dots, v_n$ are orthonormal.

Consequences:

I $v_1, \dots, v_n \in \mathbb{R}^n$ form a basis for $\mathbb{R}^n$ called the eigenbasis.

II $A = QDQ^T$ (diagonalization) $Q = (v_1 \cdots v_n)$, $D = \mathrm{diag}(\lambda_i)$.

We can use Q eigenbasis to get $x = Qw$, Q is transformation.

s.t. $x = w_1 v_1 + \dots w_n v_n$, $w = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}$

Case I: Express $f(x) = x^T A x$ in Q basis.

$$f\left(\sum_{i=1}^{n} w_i v_i\right) = \left\langle \sum_i w_i v_i, A \sum_i w_i v_i \right\rangle = \sum_i \sum_j w_i v_i A w_j v_j$$

$$= \sum_i \sum_j w_i w_j v_i \lambda_j v_j = \sum w_i^2 \lambda_i \quad \text{since } v_i \cdot v_j = 0 \ i \neq j.$$

so the cross terms are eliminated.

We can use the above to do linear approximation like Taylor Series.

Ex: Find quadratic approximation of f about (1,0) where

$$f(x_1,x_2) = x_1x_2^2 + x_1^2 + x_2^2$$

$x_0 = \begin{pmatrix}1\\0\end{pmatrix}$ $\quad \nabla f(x)\Big|_{\binom{1}{0}} = \begin{pmatrix} x_2^2 + 2x_1 \\ 2x_1x_2 + 2x_2 \end{pmatrix}\Bigg|_{\binom{1}{0}} = \begin{pmatrix}2\\0\end{pmatrix}$

$f(x_0) = 1$

so linear approx is $\ell(x) = f(x_0) + \nabla f(x_0)(x - x_0)$

$$= 1 + [2 \;\; 0]\begin{bmatrix} x_1 - 1 \\ x_2 \end{bmatrix}$$

$$= 1 + 2(x_1 - 1)$$

On top of this we use Hessian to do quadratic approx.

$$f(x) \approx f(x_0) + \nabla f(x_0)^T(x - x_0) + \frac{1}{2}(x - x_0)^T \nabla^2 f(x_0)(x - x_0).$$

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}, \quad \nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & & & \vdots \\ \vdots & & & \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

so $q = \ell(x) + \nabla^2 f(x)(x - x_0)$.

$$\nabla^2 f(x)\Big|_{\binom{1}{0}} = \begin{pmatrix} 2 & 2x_2 \\ 2x_2 & 2x_1 + 2 \end{pmatrix}\Bigg|_{\binom{1}{0}} = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}.$$

so $q = 2x_1 - 1 + \frac{1}{2}\begin{pmatrix} x_1 - 1 \\ x_2 \end{pmatrix}^T \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}\begin{pmatrix} x_1 - 1 \\ x_2 \end{pmatrix} = 2x_1 - 1 + \frac{1}{2}(x_1 - 1)(2x_2 - 2) + 4x_2^2.$