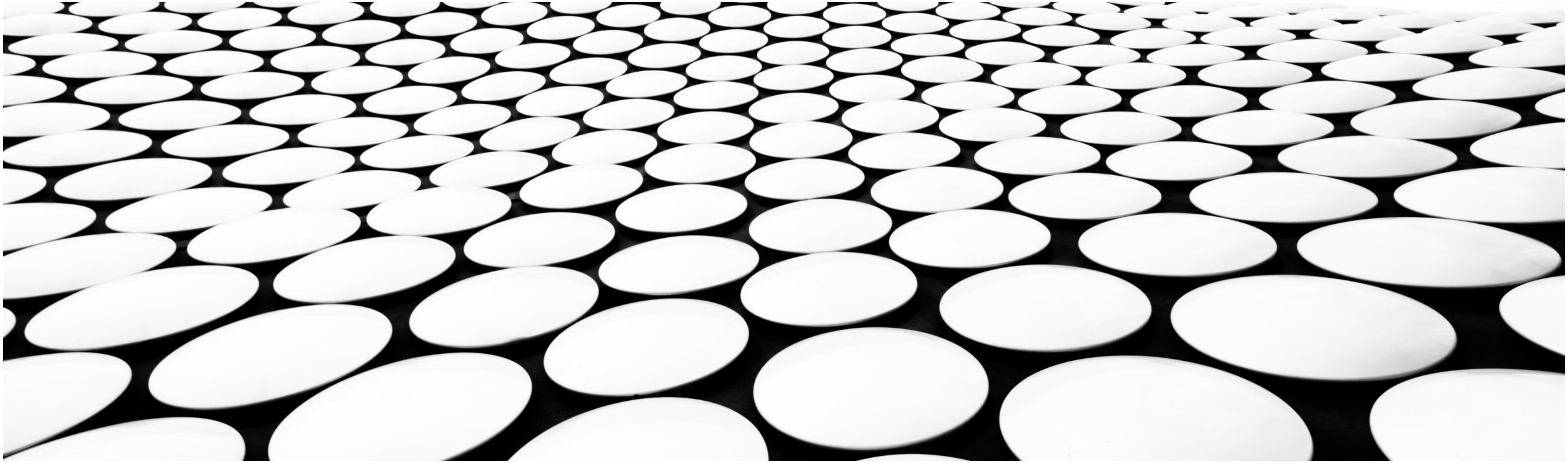
NONLINEAR OPTIMIZATION



GRADIENT

- Gradient of $f(x_1, x_2, \ldots, x_n)$: $\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$

HESSIAN MATRIX

- Hessian of $f(x_1,x_2,\ldots,x_n)$: $\mathbf{H}_f = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1\,\partial x_n} \\[2.2ex] \dfrac{\partial^2 f}{\partial x_2\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2\,\partial x_n} \\[2.2ex] \vdots & \vdots & \ddots & \vdots \\[2.2ex] \dfrac{\partial^2 f}{\partial x_n\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$

POSITIVE [SEMI]DEFINITE MATRIX

- A symmetric matrix *M* with real entries is positive definite if the real number $z^{t} * M * z > 0$ for every nonzero real column vector *z*.
- A symmetric matrix *M* with real entries is positive semidefinite if the real number $z^{t} * M * z \geq 0$ for every nonzero real column vector *z*.
- If neither A nor −A is positive semi-definite, we say that A is indefinite.

NEGATIVE DEFINITE MATRIX

- A symmetric matrix M with real entries is negative-definite if the real number $z^{t} * M * z < 0$ for every nonzero real column vector z.

MATRIX DETERMINANT

The determinant of a 2×2 matrix is

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc,$$

and the determinant of a 3×3 matrix is

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei + bfg + cdh - ceg - bdi - afh.$$

MATRIX DETERMINANT

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{vmatrix} = a_{11} * \begin{vmatrix} a_{22} & a_{23} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n2} & a_{n3} & \dots & a_{nn} \end{vmatrix} - a_{12} * \begin{vmatrix} a_{21} & a_{23} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n3} & \dots & a_{nn} \end{vmatrix} + a_{13} * \begin{vmatrix} a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} - a_{14} * \dots$$

EIGENVALUES

- To calculate the eigenvalues of a square matrix, we need to solve the characteristic equation for the matrix: $(A-\lambda I)x=0$

  x is the eigenvector associated with the eigenvalue λ.

- To find non-trivial solutions for x, the determinant of the matrix $(A-\lambda I)$ must be equal to 0.

$$\det(A-\lambda I)=0$$

- The solutions are the eigenvalues of the matrix A.
- There can be up to n eigenvalues for an n x n matrix.

EIGENVALUES

The Courant-Fischer Theorem

Let $\boldsymbol{A}$ *be a symmetric matrix in* $\mathbf{R}^{n\times n}$, *with eigenvalues* $\lambda 1 \le \lambda 2 \le \ldots \le \lambda$n. *Then*

$$\lambda_i = \min_{\substack{\textit{subspace } W \subseteq \mathbb{R}^n \\ \dim(W)=i}} \max_{\boldsymbol{x}\in W, \boldsymbol{x}\neq \boldsymbol{0}} \frac{\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{x}}$$

$$\lambda_i = \max_{\substack{\textit{subspace } W \subseteq \mathbb{R}^n \\ \dim(W)=n+1-i}} \min_{\boldsymbol{x}\in W, \boldsymbol{x}\neq \boldsymbol{0}} \frac{\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{x}}$$

Theorem

Let **A** be a symmetric matrix in $R^{n \times n}$.

1. **A** is positive definite iff all its eigenvalues are positive;
2. **A** is positive semidefinite iff all its eigenvalues are non-negative

LOCAL MINIMUM

- If $\mathbf{x}$ is a local extremum of a differentiable function f then $\nabla f(\mathbf{x}) = \mathbf{0}$.
- Let f be a function twice differentiable at $\mathbf{x}$. If $\mathbf{x}$ is a local minimum, then $\mathbf{H}f(\mathbf{x})$ is positive semidefinite.
- If $Hf(x)$ is positive definite then x is a local minimum

CONVEXITY

- If $Hf(x)$ is positive semidefinite for any $x \in S$ then f is convex on S.
- If $Hf(x)$ is positive definite for any $x \in S$ then f is strictly convex on S.
- If f is convex, then $Hf(x)$ is positive semi-definite $\forall x \in S$.

UNCONSTRAINED OPTIMIZATION

- The optimization algorithms generate a sequence $\{x_n\}$ starting from some x_0 using a displacement operation:

$$x_{n+1} \leftarrow x_n + \tau_n . d_n$$

- τ_n. is a stepsize coefficient along the direction d_n.
- Methods differentiate in the way the direction and the stepsize coefficient are chosen.

GRADIENT DESCENT

- Start from an initial position x_0
- This idea is to take a step in the direction of $-\nabla f(x_0)$
- For a one variable function:
  - $f'(x_0)>0$ when f is increasing; we should make a negative move to get towards the minimum value.
  - $f'(x_0)<0$ when f is decreasing; we should make a positive move to get towards the minimum value.
- Apply iteratively $x_{n+1} \leftarrow x_n - \tau_n . \nabla f(x_n)$ τ is a positive "learning rate"

GRADIENT DESCENT

- For a one variable function:
  - $f'(x_0) > 0$ when f is increasing; we should make a negative move to get towards the minimum value.
  - $f'(x_0) < 0$ when f is decreasing; we should make a positive move to get towards the minimum value.
- The choice of τ influences the speed of convergence.
- A small step size makes convergence too slow.
- A large step size may lead to divergence.

GRADIENT DESCENT

- Barzilai and Borwein (1988) proposed a particular rule for computing the stepsize in gradient methods that leads to a significant improvement in numerical performance. In particular the stepsize is chosen as:

$$\tau_k = \frac{\|x_k - x_{k-1}\|^2}{(x_k - x_{k-1})^T(\nabla f(x_k) - \nabla f(x_{k-1}))}$$

CONJUGATE GRADIENT METHODS

- Introduced first for minimizing strictly convex quadratic functions
- Requires only first order derivatives

$$f(x) = \frac{1}{2}(x_1, x_2, \dots, x_n).\, Q.\begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} + (a_1, a_2, \dots, a_n).\begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}$$

$$f(x) = \frac{1}{2}x^T.Q.x + a^T.x$$

with Q symmetric positive definite.

CONJUGATE DIRECTIONS ALGORITHM (CDA)

- Minimizing f can be split into n minimizations over IR.
- This is done by means of n directions $d_0, d_1, \ldots, d_{n-1}$ conjugate with respect to the Hessian Q.

$$d_j^T . Q . d_i = 0 \text{ for } i \neq j$$

  with Q symmetric positive definite.

- The global minimizer of a strictly convex quadratic function is found in at most n iterations

CONJUGATE DIRECTIONS ALGORITHM (CDA)

- Starting from x_0, $x_{k+1} \leftarrow x_k + \tau_k . d_k$

$$\tau_k = \frac{-\nabla f(x_k)^T . d_k}{d_k^T . Q . d_k}$$

$$d_0 = -\nabla f(x_0)$$

$$d_k = -\nabla f(x_k) + \beta_{k-1} d_{k-1}$$

β_{k-1} is calculated using either the Fletcher-Reeves formula:

$$\beta_{k-1} = \frac{\|\nabla f(x_k)\|^2}{\|\nabla f(x_{k-1})\|^2}$$

or the Polak-Ribiére formula:

$$\beta_{k-1} = \frac{\nabla f(x_k)^T (\nabla f(x_k) - \nabla f(x_{k-1}))}{\|\nabla f(x_{k-1})\|^2}$$

NEWTON'S METHOD

- relies on the quadratic approximation

$$x_{k+1} = x_k - \tau_k [\nabla^2 f(x_k)]^{-1}.\nabla f(x_k)$$

$\nabla^2 f(x_k)$ is the Hessian matrix