Linear Algebra

A subspace is a set $S \subseteq \mathbb{R}^n$ such that $0 \in S$ and $\forall x, y \in S, \beta \in \mathbb{R}, \alpha x + \beta y \in S$. $x \in \mathbb{R}^n$ is a linear combination of $v_1, \ldots, v_k$ if $x = \sum_{i=1}^{k} \alpha_i v_i$. The span of $\{v_1, \ldots, v_k\} \subseteq \mathbb{R}^n$ is $\text{Span}(S) = S$. 

The dimension of $S$ is $|B|$, a basis of $S$. For $S$ and $T$, $\dim(S \cap T) = \dim(S) + \dim(T) - \dim(S + T)$. 

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ has $\text{im}(T) = \mathbb{R}^m$, $\ker(T) = \{0\}$. It is invertible if and only if $\dim(\text{im}(T)) = \dim(S)$ and $\dim(\ker(T)) = 0$. 

The inverse of a matrix $A$ is $A^{-1}$, if it exists. $A^{-1}B = I$ and $BA^{-1} = I$. 

The Frobenius norm of a matrix $A$ is $\|A\|_F = \sqrt{\sum_{i,j} |a_{ij}|^2}$.

**Gaussian Elimination**

GE produces a factorization $A = LU$, GEPP (PAUL) = $AU$. 

**Plain GE**

1. for $k = 1$ to $n - 1$ do
2. $\alpha_{k,j} \leftarrow \text{gammax}(|a_{k,j}|)$
3. $k = \text{argmax}(|a_{k,j}|)$
4. $a_{k,j} \leftarrow 1$; $a_{k,k} \leftarrow \text{sqrt}(\sum_{j=1}^{k} |a_{k,j}|^2)$
5. $\beta_{k,j} = \text{sqrt}(\sum_{j=1}^{k} |a_{k,j}|^2)$
6. $\beta_{k,j} \leftarrow \beta_{k,j}$
7. for $j = 1$ to $k - 1$ do
   1. $x_{j,k} \leftarrow x_{j,k}$
   2. $x_{j,k} \leftarrow x_{j,k} - \alpha_{k,j} x_{j,k}$
   3. $x_{j,k} \leftarrow x_{j,k}$
8. for $k = 1$ to $n$ do
   1. $x_{k,k} \leftarrow x_{k,k} - \sum_{j=1}^{k} \alpha_{k,j} x_{j,k}$

**Orthogonal Matrices**

1. $Q^T Q = Q Q^T = I$
2. $\|x\|_2$ for each row and column of $Q$. The inner product of any row (or column) with another is 0.
3. $x^T x = \sum_{i=1}^{n} x_i^2$ for $x \in \mathbb{R}^n$.
4. $x^T y = \sum_{i=1}^{n} x_i y_i = \langle x, y \rangle$ for $x, y \in \mathbb{R}^n$.

**Norms**

A vector norm function $\| \cdot \| : \mathbb{R}^n \to \mathbb{R}$ satisfies:
1. $\|x\| \geq 0$, $\|x\| = 0$ if and only if $x = 0$
2. $\|\alpha x\| = |\alpha| \|x\|$ for all $\alpha \in \mathbb{R}$
3. $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in \mathbb{R}^n$

**Singular Value Decomposition**

For any $A \in \mathbb{R}^{m \times n}$, we can express $A = U \Sigma V^T$ such that $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthogonal, and $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_r)$ such that $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r \geq 0$. 

1. Matrix $2$-norm, $\|A\|_2 = \sigma_1$
2. The condition number $\kappa(A) = \|A\|_2 \|A^{-1}\|_2$ or $\kappa(A) = \|A\|_2 \|A^{-1}\|_2 = \frac{\|A\|_2}{\|A^{-1}\|_2}$. 
3. Rank $r$ approximation to $A$, $\Sigma_r = \text{diag}(\sigma_1, \ldots, \sigma_r)$. 
4. Rank determination, since $\kappa(A) = r$ equals the number of nonzero $\sigma_i$'s or in machine arithmetic, perhaps the number of singular values $\geq \text{mach}$.

**Information Retrieval & LSI**

If $w_1, \ldots, w_n$ are words, $w_i(t)$ is the (approximate) weight of term $i$ in document $d$. The corpus matrix is $A = [w_1, \ldots, w_n]^T \in \mathbb{R}^{n \times m}$. For a query $q \in \mathbb{R}^n$, rank documents according to $\sum_{i=1}^{n} w_i^T q$. 

**Eigenvalues & Eigenvectors**

For $A \in \mathbb{C}^{n \times n}$, $\lambda$ is an eigenvalue of $A$ if $A v = \lambda v$. 

**Complex Numbers**

Complex numbers are written as $z = x + iy \in \mathbb{C}$ for $-\infty < \Re(z) < \infty$. The imaginary part is $y = \text{Im}(z) = \sqrt{-z(z^*)}$. 

**Eigenvectors**

For $A \in \mathbb{R}^{n \times n}$, $\lambda = \text{det}(A)$ is an eigenvalue of $A$. 

**Simplicity**

$A$ is singular if $\text{det}(A) = 0$. 

$A$ is diagonalizable if it has a complete set of linearly independent eigenvectors. 

For symmetric $A \in \mathbb{R}^{n \times n}$, $A$ is diagonalizable iff $A$ is a real symmetric matrix. 

**Orthogonal Matrices**

For $Q \in \mathbb{R}^{n \times n}$, $Q^T Q = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$. 

**QR-factorization**

For any $A \in \mathbb{R}^{m \times n}$ with $m \geq n$, we can factor $A = QR$, where $Q \in \mathbb{R}^{m \times m}$ and $R \in \mathbb{R}^{n \times n}$. 

**QR-Factorization**

For any $A \in \mathbb{R}^{m \times n}$ with $m \geq n$, we can factor $A = QR$, where $Q \in \mathbb{R}^{m \times m}$ and $R \in \mathbb{R}^{n \times n}$. 

**QR-Factorization**

For any $A \in \mathbb{R}^{m \times n}$ with $m \geq n$, we can factor $A = QR$, where $Q \in \mathbb{R}^{m \times m}$ and $R \in \mathbb{R}^{n \times n}$. 

If $|\lambda| = \|\lambda\|_1$, we factor $A = QR$. 

**QR-Factorization**

For any $A \in \mathbb{R}^{m \times n}$ with $m \geq n$, we can factor $A = QR$, where $Q \in \mathbb{R}^{m \times m}$ and $R \in \mathbb{R}^{n \times n}$. 

**QR-Factorization**

For any $A \in \mathbb{R}^{m \times n}$ with $m \geq n$, we can factor $A = QR$, where $Q \in \mathbb{R}^{m \times m}$ and $R \in \mathbb{R}^{n \times n}$.
In the splitting method, $A = M - N$ and $Mv = c$ is easily solvable. Then, $x_{k+1} = M^{-1}(Nx_{k} + b)$. If it converges, the limit point $x^*$ is a solution to $Ax = b$.

The error is $e_k = (M^{-1}N)e_0$, so splitting methods converge if $\lambda_{\max}(-M^{-1}N) < 1$.

In the Jacobi method, consider $M$ as the diagonal of $A$. This will fail if $A$ has any zero diagonals.

**Conjugate Gradient**

Conjugate gradient iteratively solves $Ax = b$ for SPD $A$. It is derived from Lanczos and exploits that if $A$ is SPD then $T$ is SPD. It produces the exact solution after $n$ iterations. Time per iteration is $O(n)$.

For $k = 0, 1, 2, \ldots$ do
1. $r_{0} = b - Ax_{0}$
2. $p_{0} = r_{0}$
3. For $k = 0, 1, 2, \ldots$ do
   - $\alpha_k = \frac{r_{k}^T r_{k}}{p_{k}^T A p_{k}}$
   - $x_{k+1} = x_k + \alpha_k p_k$
   - $r_{k+1} = r_k - \alpha_k Ap_k$
   - $\beta_k = \frac{r_{k+1}^T r_{k+1}}{r_{k}^T r_{k}}$
   - $p_{k+1} = r_{k+1} + \beta_k p_k$
4. end for

**Multivariate Calculus**

Provided $f : \mathbb{R}^m \rightarrow \mathbb{R}$, the gradient and Hessian are

$$
\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_m} \end{bmatrix}, \quad
\nabla^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_m} \\
\vdots & \ddots & \vdots \\
\frac{\partial^2 f}{\partial x_m \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_m^2} \end{bmatrix}
$$

If $f$ is $c^2$ (2nd partials are all continuous), $\nabla^2 f$ is symmetric.

The Taylor expansion for $f$ is

$$
f(x + h) = f(x) + h^{T} \nabla f(x) + \frac{h^{T}}{2} \nabla^2 f(x)(h) + O(|h|^3)
$$

Provided $f$ is convex on $T$, the gradient is $\nabla f(x)$. The Hessian is $\nabla^2 f(x)$.

**Nonlinear Equation Solving**

Given $f : \mathbb{R}^m \rightarrow \mathbb{R}$, we want $x$ such that $f(x) = 0$.

In fixed point iteration, choose $g : \mathbb{R}^m \rightarrow \mathbb{R}^m$ such that $x^{k+1} = g(x^k)$. If it converges to $x^*$, $g(x^*) = x^*$.

A linear (or quadratic) model approximates a function $f$ by the first two (or three) terms of $f$'s Taylor expansion.

**ODE (or PDE) has one (or multiple) independent variables.**

In initial value problems, given $y(0) = y_0$, $y(t) \in \mathbb{R}$, and $y(0) = y_0$, we want $y(t)$ for $t > 0$.

1. Exponential growth/decelay with $y(t) = y_0 e^{\lambda t}$, with closed form $y(t) = y_0 e^{\lambda t}$. Growth if $\lambda > 0$, decay if $\lambda < 0$.

2. Ecological models, $y_0 = f(t_0, \ldots, t_n)$ for species $i = 1, \ldots, n$, is population size, $f$ encodes species relation- ships.

3. Mechanics, e.g. wall-spring-block models for $F = ma$ and $F = -kx$, so $\frac{dx}{dt} = \frac{F}{m}$; $\frac{dx}{dt} = kx$. Yields $\frac{dx}{dt} = -\frac{F}{m}$.

4. Stability of an ODE, let $y = Ay$ for $A \in \mathbb{C}^{n \times n}$. The stable or neutrally stable or unstable case is where, $\Re\lambda(A) < 0$ or $\Re\lambda(A) = 0$ or $\Re\lambda(A) > 0$, respectively.

5. In finite difference methods, approximate $y(t)$ by discrete points $y_0$, given $y_1, y_2, \ldots, y_n$, solve $y_0 = y(t_0)$ for increasing $t_0$.

For many IVPs and FDs, if the local truncation error (error at each step) is $O(k)$, then $O(k)$, the global truncation error (error overall) is $O(h^n)$. Call $p$ the order of accuracy.

To find $p$, substitute the exact solution into FD formula, insert a remainder term $r$ on RHS, use a Taylor series expansion, solve for $r$, keep only the leading term.

In Euler’s method, $y_{k+1} = y_k + hf(t_k, y_k)$, where $h = t_{k+1} - t_k$ is the step size, and $y_k = f(t_k, y_k)$. Euler’s method is computed by finit difference: $p = 1$, very explicit! $p = 1$.

A stiff problem has widely varying time scales in the solution, e.g., a transient initial velocity that in the true solution disappears immediately, chemical reaction rate variability over temperature, transients in electrical circuits. An explicit method requires $h_k$ to be on the smallest scale! Backward Euler has $y_{k+1} = y_k + hf(y_{k+1}, t_{k+1})$. BE is im- plicit ($y_{k+1}$ on the RHS). If the original program is stable, any $h_k$ will work.

**Miscellaneous**

$$
\begin{align*}
\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_{i} x_{j} &= \text{Pascal's Triangle} + O(n^p) \\
ax^2 + bx + c &= 0, \quad r_1, r_2 = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \\
\text{Exact arithmetic is slow, futile for inexact observations, and } \\
\text{NA relies on approximate algorithms.}
\end{align*}
$$