Appendix A

Basic facts from linear algebra

We assume that the reader is familiar with the basic notions of linear algebra.

A.1 Linear maps and linear groups

A *linear transformation* of a linear (vector) space (modeled as $\mathbb{R}^n$) is defined as a map $T : \mathbb{R}^n \to \mathbb{R}^n$ such that

- $T(x + y) = T(x) + T(y) \ \forall \ x, y \in \mathbb{R}^n$
- $T(\alpha x) = \alpha T(x) \ \forall \ x \in \mathbb{R}^n, \alpha \in \mathbb{R}$.

If we consider (the ring of) all $n \times n$ matrices over the field $\mathbb{R}$, its group of units $\mathcal{GL}(n)$ – which consists of all $n \times n$ *invertible* matrices and is called the *general linear group* – can be identified with the set of linear maps:

$$T : \mathbb{R}^n \to \mathbb{R}^n; \ x \mapsto T(x) = \mathbf{T}x \mid \mathbf{T} \in \mathcal{GL}(n). \tag{A.1}$$

We recall that a set G is a *group* if it closed with respect to an operation, call it $\cdot$

$$\begin{aligned} \cdot : G \times G &\longrightarrow G \\ (g_1, g_2) &\mapsto g_1 \cdot g_2 \end{aligned} \tag{A.2}$$

which is associative, has a null element and an inverse:

1. $(g_1 \cdot g_2) \cdot g_3 = g_1 \cdot (g_2 \cdot g_3) \ \forall \ g_1, g_2, g_3 \in G$ (associative)
2. $\exists \ e \in G \mid g \cdot e = g \ \forall g \in G$ (null element)
3. $\forall \ g \in G \ \exists \ g^{-1} \in G \mid g \cdot g^{-1} = g^{-1} \cdot g = e$ (inverse).

The set of $n \times n$ non-singular matrices is a group under the usual matrix product. Such a group can also be identified with the *metric* (vector) space $\mathbb{R}^{n^2}$.

We say that a linear transformation of a space with inner product is *orthogonal* if it preserves such inner product:

$$\langle \mathbf{T}x, \mathbf{T}y \rangle = \langle x, y \rangle \ \forall \ x, y \in \mathbb{R}^n. \tag{A.3}$$

The set of $n \times n$ orthogonal matrices forms the *orthogonal group* $O(n)$. If M is a matrix representative of an orthogonal transformation, expressed relative to an orthonormal reference frame, then it is easy to see that the orthogonal group is characterized as

$$O(n) = \{M \in \mathcal{GL}(n) \mid MM^T = I\}. \tag{A.4}$$

The determinant of an orthogonal matrix can be ± 1. The subgroup of $O(n)$ with unit determinant is called the *special orthogonal group* $SO(n)$.

A.2 Gram-Schmidt orthonormalization

A matrix in $\mathcal{GL}(n)$ has n independent rows (columns). A matrix in $O(n)$ has orthonormal rows (columns). The Gram-Schmidt procedure can be viewed as a map between $\mathcal{GL}(n)$ and $O(n)$, for it transforms a nonsingular matrix into an orthonormal one. Call $\mathcal{L}_+(n)$ the subset of $\mathcal{GL}(n)$ consisting of lower triangular matrices with positive elements along the diagonal. Such matrices form a subgroup of $\mathcal{GL}(n)$. Then we have

Theorem A.1. *(Gram-Schmidt) $\forall M \in \mathcal{GL}(n)\ \exists! L \in \mathcal{L}_+(n)\ E \in O(n)$ such that*

$$M = LE \tag{A.5}$$

Proof. The proof consists in constructing L and E iteratively from the rows $\mathbf{m}_{i.}$ of M:

$$\begin{array}{rclcl}
\mathbf{v}_{1.} & \doteq & \mathbf{m}_{1.} & \longrightarrow & \mathbf{e}_{1.} \doteq \frac{\mathbf{v}_{1.}}{\|\mathbf{v}_{1.}\|} \\
\mathbf{v}_{2.} & \doteq & \mathbf{m}_{2.} - \langle \mathbf{m}_{2.}, \mathbf{e}_{1.} \rangle \mathbf{e}_{1.} & \longrightarrow & \mathbf{e}_{2.} \doteq \frac{\mathbf{v}_{2.}}{\|\mathbf{v}_{2.}\|} \\
\vdots & \doteq & \vdots & \longrightarrow & \vdots \\
\mathbf{v}_{n.} & \doteq & \mathbf{m}_{n.} - \sum_{i=1}^{n-1} \langle \mathbf{m}_{i.}, \mathbf{e}_{i.} \rangle \mathbf{e}_{i.} & \longrightarrow & \mathbf{e}_{n.} \doteq \frac{\mathbf{v}_{n.}}{\|\mathbf{v}_{n.}\|}
\end{array}$$

Then $E = [\mathbf{e}_{1.}^T \dots \mathbf{e}_{n.}^T]^T$ and the matrix L is obtained as

$$L = \begin{bmatrix} \|\mathbf{v}_{1.}\| & 0 & \dots & 0 \\ \langle \mathbf{m}_{2.}, \mathbf{e}_{1.} \rangle & \|\mathbf{v}_{2.}\| & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \dots & \|\mathbf{v}_{n.}\| \end{bmatrix}$$

□

Remark A.1. *Gram-Schmidt's procedure has the peculiarity of being* causal, *in the sense that the k-th column of the transformed matrix depends only upon rows with index $l \leq k$ of the original matrix. The choice of the name E for the orthogonal matrix above is not random. In fact we will view the Kalman filter as a way to perform a Gram-Schmidt orthonormalization on a peculiar Hilbert space, and the outcome E of the procedure is traditionally called the* innovation.

A.3 Symmetric matrices

Definition A.1. *$Q \in \mathbb{R}^{n \times n}$ is* symmetric *iff $Q^T = Q$.*

Theorem A.2. *Q is symmetric then*

1. *Let* (v, λ) *be eigenvalue-eigenvector pairs. If* $\lambda_i \neq \lambda_j$ *then* $v_i \perp v_j$*, i.e. eigenvectors corresponding to distinct eigenvalues are orthogonal.*

2. $\exists$ *n orthonormal eigenvectors of* Q*, which form a basis for* $\mathbb{R}^n$*.*

3. $Q \geq 0$ *iff* $\lambda_i \geq 0 \forall i = 1 : n$*, i.e.* Q *is positive semi-definite iff all eigenvalues are non-negative.*

4. *if* $Q \geq 0$ *and* $\lambda_1 \geq \lambda_2 \cdots \lambda_n$ *then* $\max_{||x||_2=1}\langle x, Qx\rangle = \lambda_1$ *and* $\min_{||x||_2=1}\langle x, Qx\rangle = \lambda_n$*.*

Remark A.2.

- *from point (3) of the previous theorem we see that if* $V = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$ *is the matrix of all the eigenvectors, and* $\Lambda = diag\{\lambda_1 \cdots \lambda_n\}$ *is the diagonal matrix of the corresponding eigenvalues, then we can write* $Q = V \Lambda V^T$*; note that* V *is orthonormal.*

- *Proofs of the above claims are easy exercises.*

Definition A.2. *Let* $A \in \mathbb{R}^{m \times n}$*, then we define the* <u>*induced 2-norm*</u> *of* A *as an operator between* $\mathbb{R}^n$ *and* $\mathbb{R}^m$ *as*

$$\|A\| \doteq \max_{\|x\|_2=1} \|Ax\|_2^2 = \max_{\|x\|_2=1} \langle x, A^T A x\rangle.$$

Remark A.3.

- *Similarly other induced operator norms on* A *can be defined starting from different norms on the domain and co-domain spaces on which* A *operates.*

- *let* A *be as above, then* $A^T A$ *is clearly symmetric and positive semi-definite, so it can be diagonalized by a orthogonal matrix* V*. The eigenvalues, being non-negative, can be written as* σ_i^2*. By ordering the columns of* V *so that the eigenvalue matrix* Λ *has decreasing eigenvalues on the diagonal, we see, from point (e) of the previous theorem, that* $A^T A = V diag\{\sigma_1^2 \cdots \sigma_n^2\} V^T$ *and* $\|A\|_2 = \sigma_1$*.*

A.4 Structure induced by a linear map

- Let A be an operator from a vector space E to a space F

- Let E have a scalar product $\langle\ ,\ \rangle_E : E \times E \longrightarrow \mathbb{F}$ and F have <u>finite dimension</u> and a scalar product $\langle\ ,\ \rangle_F : F \times F \longrightarrow \mathbb{F}$

- Let E be decomposed as:

$$E = Nu(A) \overset{\perp}{\oplus} Nu(A)^{\perp}$$

- Let F be decomposed as $E = Ra(A) \overset{\perp}{\oplus} Ra(A)^{\perp}$.

Theorem A.3. *Let* A, E, F *be defined as above; then*

a) $Nu(A)^{\perp} = Ra(A^T)$

b) $Ra(A)^{\perp} = Nu(A^T)$

c) $Nu(A^T) = Nu(AA^T)$

d) $Ra(A)^{\perp} = Ra(AA^T)$*.*

A.5 The Singular Value Decomposition (SVD)

The SVD is a useful tool to capture essential features of a linear operator, such as the rank, range space, null space, induced norm etc. and to "generalize" the concept of "eigenvalue- eigenvector" pair.

The computation of the SVD is numerically well-conditioned, so it makes sense to try to solve some typical linear problems as matrix inversions, calculation of rank, best 2-norm approximations, projections and fixed-rank approximations, in terms of the SVD of the operator.

A.5.1 Algebraic derivation

Theorem A.4. *Let $A \in \mathbb{R}^{m \times n}$ have rank p. Furthermore suppose, WLOG, that $m \geq n$, then*

- $\exists U \in \mathbb{R}^{m \times p}$ *whose columns are orthonormal*
- $\exists V \in \mathbb{R}^{n \times p}$ *whose columns are orthonormal*
- $\exists \Sigma \in \mathbb{R}^{p \times p}, \Sigma = diag\{\sigma_1 \cdots \sigma_p\}$ *diagonal with* $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p$

such that $A = U \Sigma V^T$.

Constructive proof

- compute $A^T A$: it is symmetric and positive semi-definite of dimension $n \times n$. Then order its eigenvalues in decreasing order and call them $\sigma_1^2 \geq \cdots \geq \sigma_p^2 \geq \cdots \sigma_n^2 \geq 0$. Call the σ_i *singular values.*
- From an orthonormal set of eigenvectors of $A^T A$ create an orthonormal basis for $\mathbb{R}^n$ such that $span\{v_1 \cdots v_p\} = Ra(A^T)$ and $span\{v_{p+1} \cdots v_n\} = Nu(A)$. Note that the latter eigenvectors correspond to the zero singular values, since $Nu(A^T A) = Nu(A)$.
- define u_i such that $Av_i = \sigma_i u_i \forall i = 1 : p$, and see that the set $\{u_i\}$ is orthonormal (proof left as exercise).
- Complete the basis $\{u_i\}_{\forall i=1:p}$, which spans $Ra(A)$ (by construction), to all $\mathbb{R}^m$.
- then $A\,[v_1 \cdots v_n] = [u_1 \cdots u_m] \begin{bmatrix} \sigma_1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \sigma_2 & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \sigma_p & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \vdots & \vdots & 0_n \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & \cdots & 0_m \end{bmatrix}$ which we name $A\tilde{V} = \tilde{U}\tilde{\Sigma}$
- hence $A = \tilde{U}\tilde{\Sigma}\tilde{V}^T$

Then the claim follows by deleting the columns of $\tilde{U}$ and the rows of $\tilde{V}^T$ which multiply the zero singular values.

A.5.2 Geometric interpretation

Theorem A.5. *Let $A \in \mathbb{R}^{n \times n} = U\Sigma V^T$, then A maps $B(0,1) \doteq \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$ to an ellipsoid with half-axes $\sigma_i u_i$*

Proof:

let x, y be such that $Ax = y$. $\{v_1 \cdots v_n\}$ is an orthonormal basis for $\mathbb{R}^n$. With respect to such basis x has coordinates $[\langle v_1, x\rangle, \langle v_2, x\rangle, \cdots, \langle v_n, x\rangle]$. Idem for $\{u_i\}$. Let $y = \sum_{i=1}^n y_i u_i \rightarrow Ax = \sum_{i=1}^n \sigma_i u_i v_i^T x = \sum_{i=1}^n \sigma_i u_i \langle v_i, x\rangle = \sum_{i=1}^n y_i u_i = y$. Hence $\sigma_i \langle v_i, x\rangle = y_i$. Now $\|x\|_2^2 = \sum_{i=1}^n \langle v_i, x\rangle^2 = 1 \forall x \in B(0,1)$, from which we conclude $\sum_{i=1}^n \frac{y_i^2}{\sigma_i^2} = 1$, which represents the equation of an ellipsoid with half-axes of length σ_i.

A.5.3 Some properties of the SVD

Rank and Null space

Theorem A.6. *Let$A = U\Sigma V^T$ have rank r; then*

- $Nu(A) = \text{span}\{v_{r+1} \dots v_n\}$
- $Ra(A^T) = Nu(A)^\perp = \text{span}\{v_1 \dots v_r\}$
- $Ra(A) = \text{span}\{u_1 \dots u_r\}$
- $Ra(A)^\perp = Nu(A^T) = \text{span}\{u_{r+1} \dots u_n\}$

proof: by construction.

Generalized (Moore-Penrose) Inverse

The problems involving orthogonal projections onto invariant subspaces of A, as Linear Least Squares (LLSE) or Minimum Energy problems, are easily solved using the SVD.

Definition A.3. *Let $A \in \mathbb{R}^{m \times n}$, $A = U\Lambda V^T$ where Λ is the diagonal matrix with diagonal elements $(\lambda_1, \dots \lambda_r, 0 \dots 0)$; then*

$$A^\dagger = U\Lambda_{(r)}^{-1} V^T, \quad \Lambda_{(r)}^{-1} = diag(\lambda_1^{-1}, \dots \lambda_r^{-1}, 0 \dots 0)$$

Theorem A.7.

- $AA^\dagger A = A$
- $A^\dagger A A^\dagger = A^\dagger$

Least squares solution of a linear systems

Theorem A.8. *Consider the problem $Ax = b$ with $A \in \mathbb{R}^{m \times n}$ of rank $p \leq min(m, n)$, then the solution $\hat{x}$ that minimizes $\|A\hat{x} - b\|$ is given by $\hat{x} = A^\dagger b$.*

Fixed rank approximations

One of the most important properties of the SVD has to deal with fixed-rank approximations of a given operator. Given A as an operator from a space X to a space Y of rank n, we want to find an operator B from the same spaces such that it has rank $p < n$ fixed and $\|A - B\|_F$ is minimal, where the F indicates the Frobenius norm (in this context it is the sum of the singular values).

If we had the usual 2-norm and we calculate the SVD of $A = U\Sigma V^T$, then by simply setting all the singular values but the first p to zero, we have an operator $B \doteq U\Sigma_{(p)}V^T$, where $\Sigma_{(p)}$ denotes a matrix obtained from Σ by setting to zero the elements on the diagonal after the p^{th}, which has exactly the same two norm of A and satisfies the requirement on the rank.

It is not difficult to see the following result

Theorem A.9. *Let A, B be defined as above, then $\|A - B\|_F = \sigma_{p+1}$. Furthermore such norm is the minimum achievable.*

Proof: easy exercise; follows directly from the orthogonal projection theorem and the properties of the SVD given above.

Perturbations

Consider a non-singular matrix $A \in \mathbb{R}^{n \times n}$ (if A is singular substitute its inverse by the moore-penrose pseudo-inverse). Let δA be a full-rank perturbation. Then

- $|\sigma_k(A + \delta A) - \sigma_k(A)| \leq \sigma_1(\delta A)\ \forall k = 1 : n$
- $\sigma_n(A\delta A) \geq \sigma_n(A)\sigma_n(\delta A)$
- $\sigma_1(A^{-1}) = \frac{1}{\sigma_n(A)}$

Condition number

Consider again the problem $Ax = b$, and consider a "perturbed" full rank problem $(a+\delta A)x = b+\delta b$. Since $Ax = b$, then to first order approximation $\delta x = -A^{\dagger}\delta Ax$. Hence $\|\delta x\| \leq \|A^{\dagger}\|\|\delta A\|\|x\|$, from which $\frac{\|\delta x\|}{\|x\|} = \|A^{\dagger}\|\|A\|\frac{\|\delta A\|}{\|A\|} \doteq k(A)\frac{\|\delta A\|}{\|A\|}$. "$k(A)$" is called the condition number of A. It easy to see that $k(A) = \frac{\sigma_1}{\sigma_n}$.