section

1 Fundations

This sections focuses on important mathematical entities that are important for formulating, analyzing the Conjugate Gradient and the Lanczos Algorithm. Major parts of this sections cited from...

1.1 Projectors

There are 2 types of projector, an oblique Projector and Orthgonal Projector. A matrix P is called a projector if:

Definition 1.

$$P^2 = P \tag{1.1.1}$$

This property is sometimes referred as idempotent. As a consequence, $\text{ran}(I-P) = \text{null}(P)$ and here is the proof:

Proof.

$$\forall x \in \mathbb{C}^n : P(I-P)x = \mathbf{0} \implies \text{ran}(I-P) \subseteq \text{null}(P) \tag{1.1.2}$$

$$\forall x \in \text{null}(P) : Px = \mathbf{0} \implies (I-P)x = x \implies x \in \text{ran}(I-P) \tag{1.1.3}$$

$$\implies \text{ran}(I-P) = \text{null}(P) \tag{1.1.4}$$

□

This consequence states the fact that any vector x can be represented in the form of: $x = Px + (I-P)x$, and every projector will be defined via the range of $I-P$ and P.

1.1.1 Orthogonal Projector

An orthogonal projector is a projector such that:

Definition 2.

$$\text{null}(P) \perp \text{ran}(P) \tag{1.1.5}$$

This property is in fact, very special. A good example of an orthogonal projector would be the Householder Reflector Matrix. Or just any $\hat{u}\hat{u}^H$ where $\hat{u}$ is being an unitary vector. For convenience of proving, assume subspace $M = \text{ran}(P)$. Consider the following lemma:

Lemma 1.

$$\text{null}(P^H) = \text{ran}(P)^\perp \tag{1.1.6}$$

$$\text{null}(P) = \text{ran}(P^H)^\perp \tag{1.1.7}$$

Using (1.1.4) and consider the proof:

Proof.

$$\langle P^H x, y\rangle = \langle x, Py\rangle \tag{1.1.8}$$

$$\forall x \in \text{null}(P^H), y \in \mathbb{C}^n \tag{1.1.9}$$

$$\implies \langle P^H x, y\rangle = 0 = \langle x, Py\rangle \tag{1.1.10}$$

$$\implies \text{null}(P^H) \perp \text{ran}(P) \tag{1.1.11}$$

$$\forall y \in \text{null}(P), x \in \mathbb{C}^n : \tag{1.1.12}$$

$$\langle x, Py\rangle = 0 = \langle P^H x, y\rangle \tag{1.1.13}$$

$$\implies \text{ran}(P^H) \perp \text{null}(P) \tag{1.1.14}$$

□

Proposition 1. A projector is orthogonal iff it's Hermitian.

Proof. $\Longleftarrow$ Assuming the matrix is Hermitian and it's a projector, then we wish to prove that it's an orthogonal projector. Let's recall:

$$\text{null}(P^H) = \text{ran}(P)^\perp \tag{1.1.15}$$

$$\text{null}(P) = \text{ran}(P^H)^\perp \tag{1.1.16}$$

Substituting $P^H = P$, we have $\text{null}(P) = \text{ran}(P)^\perp$, Which is the definition of Orthogonal Projector. Therefore, P is an orthogonal projector by the definition of the projector.

For the $\Longrightarrow$ direction, we assume that P is an Orthogonal Projector, then we wish to show that it's also Hermitian. Observe that P^H is also a projector because $(P^H)^2 = (P^2)^H$. Then, using the definition of orthogonal projector:

$$\text{null}(P) \perp \text{ran}(P) \tag{1.1.17}$$

$$\text{null}(P^H) \perp \text{ran}(P^H) \tag{1.1.18}$$

Notice that using above statement together with Lemma 1 means $\text{null}(P) = \text{ran}(P)^\perp = \text{null}(P^H)$, and then $\text{ran}(P) = \text{null}(P)^\perp = \text{ran}(P^H)$. Therefore, P^H is an projector such that: $\text{ran}(P) = \text{ran}(P^H) \wedge \text{null}(P) = \text{null}(P^H)$. The range and null space of P^H and P is the same therefore P has to be Hermitian. $\square$

1.1.2 Oblique Projector

An oblique projector is not orthogonal, and vice versa. It's a projector that satisfies the following conditions:

Definition 3.

$$Px \in M \quad (I - P)x \perp L \quad \text{where: } M \neq L \tag{1.1.19}$$

An orthgonal projector is the case when the subspace $M = L$.

A famous example of an orthogonal projector is QQ^H where Q is an Unitary Matrix. This is a Hermitian Matrix and it's idempotent, making it an orthogonal projector.

1.1.3 Projector Geometric Intuitions

A projector describes a given vector using some elements from another basis. The oblique projector creates a light sources in the form of the subspace L and it shoots parrell light ray orthogonal to L, crossing vectors and projecting their shadow onto subspace M.

1.2 Projectors and Norm Minimizations

An orthogonal projector always reduce the 2 norm of a vector. Given any subspace M, we can create a basis of vectors packing into the some matrix, say A, then P_M as a projector onto the basis M one example can be: $A(AA^T)^{-1}A^T$. Let's consider the claim:

$$\|P_M x\|^2 \leq \|x\|^2 \tag{1.2.1}$$

Proof:

$$x = Px + (I - P)x \tag{1.2.2}$$

$$\|x\|^2 = \|Px\|^2 + \|(I - P)x\|^2 \tag{1.2.3}$$

$$\|x\|^2 \geq \|Px\|^2 \tag{1.2.4}$$

Using this property of the Orthogonal Projector, we consider the following minimizations problem:

$$\min_{x \in M} \|y - x\|_2^2 = \|y - P_M(y)\|_2^2 \tag{1.2.5}$$

Proof:

$$\|y - x\|_2^2 = \|y - P_M y + P_M y - x\|_2^2 \tag{1.2.6}$$
$$\|y - x\|_2^2 = \|y - P_M y\|_2^2 + \|P_M y - x\|_2^2 \tag{1.2.7}$$
$$\implies \|y - P_M y\|_2^2 \le \|y - x\|_2^2 \tag{1.2.8}$$

That concludes the proof. Observe that, $y - P_M y \perp M$ and $P_M y - x \in M$ because $P_M y, x \in M$, which allows us to split the norm of $y - x$ into 2 components. In addition using the fact that the projector is orthogonal. That concludes the proof.

1.3 Subspace Orthogonality Framework

Let $\mathcal{K}, \mathcal{L}$ be subspaces where candidates soltuions are chosen and residuals are orthogonalized against. Under the idea case the 2 subspaces spans all dimensions, and it's able to approximate all solutions and forcing the residual vector $(b - Ax)$ to be zero. This is a description of this framework:

$$\tilde{x} \in x_0 + \mathcal{K} \text{ s.t: } b - A\tilde{x} \perp \mathcal{L} \tag{1.3.1}$$

it looks for an x in the affine linear subspace $\mathcal{K}$ such that it's perpendicular to the subspace $\mathcal{L}$, or, equivalently, minimizing the projection onto the subspace $\mathcal{L}$. One interpretation of it is an projection of residual onto the basis that is orthogonal to $\mathcal{L}$.

Sometimes, for convenience and the exposition and exposing hidden connections between ideas, the above conditions can be expressed using matrix.

$$\text{Let } V \in \mathbb{C}^{n\times m} \text{ be a aasis for: } \mathcal{K} \tag{1.3.2}$$
$$\text{Let } W \in \mathbb{C}^{n\times m} \text{ be a basis for: } \mathcal{L} \tag{1.3.3}$$

We can then make us of (1.3.1) and express it in the form of:

$$\tilde{x} = x^{(0)} + Vy \tag{1.3.4}$$
$$b - A\tilde{x} \perp (\text{span} \leftarrow \text{col})(W) \tag{1.3.5}$$
$$W^T(b - A\tilde{x} - AVy) = \mathbf{0} \tag{1.3.6}$$
$$W^T r^{(0)} - W^T AVy = \mathbf{0} \tag{1.3.7}$$
$$W^T AVy = W^T r^{(0)} \tag{1.3.8}$$

And from here, we can define a simple prototype algorithm using this framworks.

$$\begin{aligned} &\text{While not converging :} \\ &\quad \text{Increase Span for: } \mathcal{K}, \mathcal{L} \\ &\quad \text{Choose: } V, W \text{ for } \mathcal{K}, \mathcal{L} \\ &\quad r := b - Ax \\ &\quad y := (W^T AV)^{-1} W^T r \\ &\quad x := x + Vy \end{aligned} \tag{6}$$

Each time, we increase the span of the subspace $\mathcal{K}, \mathcal{L}$, which gives us more space to choose the solution x, and more space to reduce the residual vector r. This idea is incredibly flexible, and we will see in later part where it reduces to a more concrete algorithm.

1.4 Subspace Minimizations Framework

Other times, iterative method will choose to build up a subspace for each step with a subspace generator, and build up the solution on this expanding subspace, but with the additional objective of minimizing the residual under certain norm. Assuming that the vector $x \in x_0 + \mathcal{K}$, and we want to minimize the residual under a norm induced by positive definite operator B. Let it be the case that the columns of matrix K span subspace $\mathcal{K}$ with $\dim(\mathcal{K}) = k$.

$$\begin{aligned} &\min_{x \in x_0 + \mathcal{K}} \|b - Ax\|_B^2 && (1.4.1)\\ &= \min_{w \in \mathbb{R}^k} \|b - A(x_0 + Kw)\|_B^2 && (1.4.2)\\ &= \min_{w \in \mathbb{R}^k} \|r_0 - AKw\|_B^2 && (1.4.3) \end{aligned}$$

We take the derivative of it and set the derivative to zero, this translate the problem to a projection problem under the A norm.

$$\begin{aligned} \nabla_w \left[\|r_0 - AKx\|_B^2\right] &= \mathbf{0} && (1.4.4)\\ (AK)^T B(r_0 - AKx) &= \mathbf{0} && (1.4.5)\\ (AK)^T Br_0 - (AK)^T BAKx &= \mathbf{0} && (1.4.6)\\ (AK)^T Br_0 &= (AK)^T BAKx && (1.4.7) \end{aligned}$$

The above formulation is tremendously powerful. I used gradient instead of projector for the simplicity of the argument. One can derive the same using projector but the math is bit more hedious.

1.5 Krylov Subspace

A Krylov Subspace is a sequence basis paramaterized by A, an linear operator, v an initial vector, and k, which basis in the sequece of basis that we are looking at.

Definition 4 (Krylov subspace)**.**

$$\mathcal{K}_k(A|b) = \text{span}(b, Ab, A^2b, \cdots A^{k-1}b) \tag{1.5.1}$$

Please immediately observe that from the definition we have:

$$\forall v : \mathcal{K}_1(A|v) \subseteq \mathcal{K}_2(A|v) \subseteq \mathcal{K}_3(A|v) \cdots \tag{1.5.2}$$

Please also observe that, every element inside of krylov subspace generated by matrix A, and an initial veoctr v can be represented as a polynomial of matrix A multiplied by the vector v and vice versa.

$$\forall x \in \mathcal{K}_j(A|v) \; \exists \; w : p_k(A|w)v = x \tag{1.5.3}$$

We use $p_k(A|w)$ to denotes a matrix polynomial with coefficients w, where w is a vector. No proof this is trivial. Take note that, we can change the field of where the scalar w is coming from, but for discussion below, $\mathbb{R}, \mathbb{C}$ doesn't matter and won't change the results.

$$p_k(A|w)v = \sum_{j=0}^{k-1} w_j A^j v \tag{1.5.4}$$

1.6 Useful Theorems

1.6.1 Cauchy Interlace Theorem

1.6.2 Caley Hamilton's Theorem

1.7 Deriving Conjugate Gradient from First Principles