Apuntes: Cauchy–Schwarz, Formas Cuadráticas y Proyección Ortogonal

1. La forma cuadrática asociada

Sea $(H, \langle \cdot, \cdot \rangle)$ un espacio de Hilbert. Para $f, g \in H$ definimos

$$Q(\lambda) \;=\; \|f - \lambda g\|^2, \qquad \lambda \in \mathbb{C}.$$

Expandiendo con el producto interno:

$$Q(\lambda) \;=\; \|f\|^2 - 2\Re\big(\lambda \langle f, g\rangle\big) + |\lambda|^2 \|g\|^2.$$

Es decir, $Q(\lambda)$ es una *parábola* en la variable λ.

2. Optimización

Como $Q(\lambda) \geq 0$ para todo λ, el valor mínimo se alcanza en

$$\lambda^* \;=\; \frac{\langle f, g\rangle}{\|g\|^2}.$$

Sustituyendo:

$$Q(\lambda^*) \;=\; \|f\|^2 - \frac{|\langle f, g\rangle|^2}{\|g\|^2}.$$

3. Desigualdad de Cauchy–Schwarz

De la positividad se sigue:

$$Q(\lambda^*) \;\geq\; 0 \quad \Longrightarrow \quad |\langle f, g\rangle|^2 \;\leq\; \|f\|^2 \, \|g\|^2.$$

Esto es precisamente la **desigualdad de Cauchy–Schwarz**.

4. Interpretación geométrica

- $Q(\lambda)$ representa la **distancia al cuadrado** entre f y el subespacio generado por g:

$$Q(\lambda) = \|f - \lambda g\|^2.$$

- El mínimo $Q(\lambda^*)$ corresponde a la **distancia ortogonal** de f al subespacio $\operatorname{span}\{g\}$.
- El vector proyectado es

$$P_g(f) = \lambda^* g = \frac{\langle f, g\rangle}{\|g\|^2}\, g.$$

- El residuo $f - P_g(f)$ es ortogonal a g:

$$\langle f - P_g(f),\, g\rangle = 0.$$

5. Continuidad

La función $Q(\lambda)$ es cuadrática y por tanto continua en λ. Esto garantiza que el paso al mínimo no requiere argumentos de límite ε–δ, sino que la geometría misma del espacio de Hilbert asegura la continuidad y la convergencia.

6. Aplicaciones en Machine Learning

La desigualdad de Cauchy–Schwarz y la interpretación de la forma cuadrática aparecen en múltiples áreas de Machine Learning:

6.1 Similitud de coseno

Para embeddings $x, y \in \mathbb{R}^n$, se define

$$\cos(\theta) = \frac{\langle x, y\rangle}{\|x\|\|y\|}.$$

Cauchy–Schwarz garantiza que $-1 \le \cos(\theta) \le 1$, haciendo válida la noción de similitud usada en NLP, visión y sistemas de recomendación.

6.2 Regularización y estabilidad

En problemas de optimización se minimiza

$$\min_w L(w) + \lambda\|w\|^2,$$

donde la norma proviene de un producto interno. Cauchy–Schwarz asegura continuidad y estabilidad de los algoritmos de gradiente.

6.3 Kernels y SVMs

Si $K(x, y)$ es un kernel en un espacio de Hilbert reproducing (RKHS),

$$|K(x, y)| \leq \sqrt{K(x, x)}\,\sqrt{K(y, y)}.$$

Esto valida el kernel como medida de similitud y sustenta métodos como SVMs y Gaussian Processes.

6.4 Acotación de errores

Para variables aleatorias X, Y,

$$|\mathbb{E}[XY]| \leq \sqrt{\mathbb{E}[X^2]}\,\sqrt{\mathbb{E}[Y^2]}.$$

Este bound se usa en análisis de varianza, consistencia de estimadores y teoría de generalización (PAC learning).

6.5 Gradiente y optimización

En descenso de gradiente,

$$|\langle \nabla f(w), d \rangle| \leq \|\nabla f(w)\| \cdot \|d\|.$$

Esto garantiza que el gradiente actúa como un funcional lineal continuo y que el paso en dirección d está controlado.

6.6 PCA y reducción de dimensionalidad

En PCA se maximiza

$$\max_{\|v\|=1} v^T \Sigma v,$$

donde Σ es la matriz de covarianza. El máximo se alcanza en un autovector principal gracias a Cauchy–Schwarz, lo que fundamenta la reducción de dimensionalidad.

Conclusión: La desigualdad de Cauchy–Schwarz puede verse como la afirmación de que la parábola

$$Q(\lambda) = \|f - \lambda g\|^2$$

nunca desciende bajo el eje horizontal. De esta manera, Cauchy–Schwarz no sólo es una desigualdad algebraica, sino también una expresión de la *continuidad geométrica* de la proyección ortogonal.