332:541 Stochastic Signals and Systems Fall
Convexity

Some of you asked for a few notes on convexity. Here they are.

Definition: A vector-argument, real valued function $g(x)$ is strictly convex iff for $\lambda \in [0,1]$ and x_1, x_2 in the domain of $g()$ we have

$$g(\lambda x_1 + (1-\lambda)x_2) \leq \lambda g(x_1) + (1-\lambda)g(x_2)$$

with equality iff $\lambda = 0, 1$.

Simple convexity (not strict) relaxes relaxes the strict inequality except at the endpoints. That is, the expression can be satisfied with equality other than at $\lambda = 0, 1$.

The above definition is powerful since it allows us to apply convexity to multivariate functions. The geometric interpretation is that for a function to be convex, it must lie below a line drawn between ANY two points in the domain of the function.

One can also use the same basic idea to define *convex sets*. For example, a set is called convex if the line connecting any two points in the set is also completely contained in the set – that is, all points on the line are also in the set for any two chosen endpoints. This concept is useful in optimization — something we do a lot of as EE's.

In any case, our definition of convexity is completely general and our old baby definition for single-variable functions is included in our super definition. Here's why. For any function $f(x)$ on some simply-connected region (x_1, x_2) (OOOOOOOH! here's another use for convex regions – all convex regions MUST be simply connected since if they're not, you can draw a line from one of the regions to another and the line will not be completely contained in the set!) we have

$$f(x) = f(a) + \frac{df(a)}{dx}(x-a) + \frac{d^2 f(\xi)}{dx^2}\frac{1}{2}(x-a)^2$$

where ξ is between a and x. This is an often forgotten fact from Calculus 101. In any case, we first see if for $\frac{d^2 f(\xi)}{dx^2} > 0$ we satisfy our expression for convexity with $a \in (x_1, x_2)$. So we let $a = \lambda x_1 + (1-\lambda)x_2$ to obtain

$$f(x_1) > f(\lambda x_1 + (1-\lambda)x_2) + f'(\lambda x_1 + (1-\lambda)x_2)\left[(1-\lambda)(x_1 - x_2)\right]$$

where the strict inequality is owed to the positivity of the second derivative. Similarly.

$$f(x_2) > f(\lambda x_1 + (1-\lambda)x_2) + f'(\lambda x_1 + (1-\lambda)x_2)\left[\lambda(x_2 - x_1)\right]$$

From these we obtain

$$\lambda f(x_1) + (1-\lambda)f(x_2) > f(\lambda x_1 + (1-\lambda)x_2)$$

Now for the reverse arrow we'd like to show

$$\{\lambda f(x_1) + (1-\lambda)f(x_2) > f(\lambda x_1 + (1-\lambda)x_2)\} \Rightarrow \left\{\frac{d^2 f}{dx^2} > 0\right\}$$

Well, I'll leave it to you to show that if there exists a single value χ for which $\frac{d^2 f(\chi)}{dx^2} < 0$ then the formal "super-convexity" definition will not be satisfied. That is, you should find it relatively easy to show that for such a χ we will have

$$\left\{\frac{d^2 f(\chi)}{dx^2} \leq 0\right\} \Rightarrow \{\lambda f(x_1) + (1-\lambda)f(x_2) \leq f(\lambda x_1 + (1-\lambda)x_2)\}$$

for some values of x_1 and x_2 and $\lambda \neq 0, 1$.