Thursday: 9/8

What about something a little harder? Let's compute:

$$D\|y-Ac\|_2^2 = D(y-Ac)^T(y-Ac)$$

Also review MVTE
show how to do it using PDs

Big-O

$f(x) = O(g(x))$ as $x \to a$

$\Leftrightarrow \exists \delta > 0 \, \exists M > 0$

st $0 < |x-a| < \delta$

$\Rightarrow |f(x)| \le M g(x)$

or $\lim_{x \to a} \frac{f(x)}{g(x)} < \infty$

OK:

$$(y - A(c+\delta c))^T (y - A(c-\delta c))$$

$$= y^T y - y^T A(c-\delta c) - [A(c-\delta c)]^T y + [A(c-\delta c)]^T [A(c-\delta c)]$$

$$= y^T y - y^T Ac - (Ac)^T y + (Ac)^T(Ac)$$
$$+ y^T A\delta c + (A\delta c)^T y - (Ac)^T A\delta c - (A\delta c)^T Ac$$
$$+ (A\delta c)^T A\delta c$$

$$= (y-Ac)^T(y-Ac) + 2y^T A\delta c - 2(Ac)^T A\delta c$$
$$+ (A\delta c)^T A\delta c$$

$$= \underbrace{\|y-Ac\|_2^2}_{\text{"}f(c)\text{"}} + \underbrace{2(y-Ac)^T A}_{D\|y-Ac\|_2^2} \delta c + O(\|\delta c\|_2^2)$$

transpose

$$\Rightarrow \nabla_c \|y-Ac\|_2^2 = 2A^T(y-Ac)$$

<u>Problem</u>: compute $\nabla f(x)$, where $x \in \mathbb{R}^n$, and

$$f(x) = \frac{1}{2}x^T A x + b^T x + c$$

- These were examples of <u>linear</u> least squares.

- Why linear? We fit a function which is a <u>linear model</u>, i.e.,

$$f(x, c) = c_0 f_0(x) + c_1 f_1(x) + \cdots + c_m f_m(x).$$

The functions f_i may be nonlinear, but they are a linear combination. Hence, fitting reduces to solving a nonlinear system.

- We can also have a nonlinear dependence on the model parameters, e.g.

$$f(x, c) = \frac{c_0 + c_1 x}{1 + c_2 x}$$

- This leads to <u>nonlinear least squares</u>:

$$\text{minimize } \| y - f(c) \|_2^2 = \sum_{i=0}^{m} (y_i - f(x_i, c))^2.$$

- How to solve? Let $F(c) = \| y - f(c) \|_2^2$. Take gradient, set equal to zero, solve. Let's see:

$$DF(c) = D \| y - f(c) \|_2^2 = D (y - f(c))^T (y - f(c))$$

$\mathbb{R}^{(m+1) \times 1}$

$$= D (y^T y - 2 f(c)^T y + f(c)^T f(c))$$

$$= 0 - 2 Df(c)^T y + 2 Df(c)^T f(c)$$

$$= -2 Df(c)^T (y - f(c)).$$

Ok... we have:

$$Df(c)^T(y - f(c)) = 0 \Rightarrow Df(c)^T y = Df(c)^T f(c).$$

Not totally clear how to proceed...

Let's try the following trick: Let's introduce an iteration

$$(c^{(0)}, c^{(1)}, \dots)$$

where $c^{(n+1)} = c^{(n)} + \Delta c^{(n)}$

Let's apply our equation at $c^{(n)}$ to get:

$$Df(c^{(n+1)})^T y = Df(c^{(n+1)})^T f(c^{(n+1)}).$$

$$= Df(c^{(n+1)})^T f(c^{(n)} + \Delta c^{(n)})$$

$$= Df(c^{(n+1)})^T \left[f(c^{(n)}) + Df(c^{(n)})\Delta c^{(n)} + O(\|\Delta c^{(n)}\|_2^2) \right]$$

Let's assume c^* is the true solution to the nonlinear least squares problem and that $c^{(n)} \to c^*$ as $n \to \infty$.

This implies that $\|c^{(n+1)} - c^{(n)}\|_2^2 = \|\Delta c^{(n)}\|_2^2 \to 0$ as $n \to \infty$, Hence, we are justified in ignoring the quadratic term " $O(\|\Delta c^{(n)}\|_2^2)$ and consider instead: (roughly):

$$Df(c^{(n)})^T y \approx Df(c^{(n)})^T \left[f(c^{(n)}) + Df(c^{(n)})\Delta c^{(n)} \right]$$

$$\Rightarrow \quad \Delta c^{(n)} \approx \left[Df(c^{(n)})^T Df(c^{(n)}) \right]^{-1} Df(c^{(n)})^T (y - f(c^{(n)})$$