18.335 Problem Set 5

Due Monday, 11 May.

Problem 1: Convexity

Recall from class that a *convex function* is a function $f(x)$ such that $f(\alpha x+(1-\alpha)y) \le \alpha f(x)+(1-\alpha)f(y)$, for $\alpha \in [0,1]$, where $x \in \mathbb{R}^n$. The inequality is $\ge$ for *concave* functions, and = for *affine* functions. A *convex set* $X \subseteq \mathbb{R}^n$ satisfies the property that if x and y are in X, then so is the line connecting x and y, i.e. so is $\alpha x+(1-\alpha)y$ for $\alpha \in [0,1]$.

(a) Show that the feasible region satisfying m constraints $f_i(x) \le 0$, $i = 1,\ldots,m$, is a convex set if the constraint functions are convex.

(b) Suppose that the feasible region is convex, but a constraint f_i is *not* convex. Why might this make optimization harder?

Problem 2: Adjoints

Consider a recurrence relation $x^n = f(x^{n-1}, p, n) \in \mathbb{R}^K$ with initial condition $x^0 = b(p)$ and P parameters $p \in \mathbb{R}^P$, as in the notes from class (see the handout on the web page). In class, and in the handout, we used the adjoint method to derive an expression for the derivatives $\frac{dg^N}{dx}$ of a function $g(x^n, p, n) \triangleq g^n$ evaluated after N steps.

In this problem, suppose that instead we want the derivative of a function G that depends on the values of x^n from *every* $n \in \{0,1,\ldots,N\}$ as follows:

$$G(p,N) = \sum_{n=0}^{N} g(x^n, p, n)$$

for some function g.

(a) One could simply use the adjoint formula from class to obtain $\frac{dG}{dp} = \sum_n \frac{dg^n}{dp}$. Explain why this is inefficient.

(b) Describe an efficient adjoint method to compute $\frac{dG}{dp}$. (Hint: modify the recurrence relation for λ^n from class to compute $\sum_n \frac{dg^n}{dp}$ via the results of a *single* recurrence.)

(c) Apply this to the example 2×2 recurrence and g function from the notes, and implement your adjoint method in Julia. Check your derivative $\frac{dG}{dp}$ with $N = 5$ against the center-difference approximation $\frac{dG}{dp_i} = [G(p_i+\delta) - G(p_i-\delta)]/2\delta$ for $p = (1,2,3,4,5)^T$ and $\delta p = 10^{-5}$.

Problem 3: BFGS

In class, we covered BFGS updates to the approximate Hessian $H^{(n)}$ from the n-th optimization step for minimizing a function $f(x)$. Let x^n be the guess for $x \in R^N$ on the n-th step. Denote $f^n = f(x^n)$, $g^n = \nabla f|_{x^n}$ as in class. The BFGS update is:

$$H^{(n+1)} = H^{(n)} + \frac{\gamma\gamma^T}{\gamma^T\delta} - \frac{H^{(n)}\delta\delta^T H^{(n)}}{\delta^T H^{(n)}\delta},$$

where $\delta = x^{n+1} - x^n$ and $\gamma = g^{n+1} - g^n$. Equivalently, since this is a pair of rank-1 updates, we can invert it via the Sherman–Morrison formula to get an update for $\left[H^{(n)}\right]^{-1}$:

$$\left[H^{(n+1)}\right]^{-1} = \left[H^{(n)}\right]^{-1} + \frac{1}{\gamma^T\delta}\left\{\left(1 + \frac{\gamma^T\left[H^{(n)}\right]^{-1}\gamma}{\gamma^T\delta}\right)\delta\delta^T - \left[H^{(n)}\right]^{-1}\gamma\delta^T - \delta\gamma^T\left[H^{(n)}\right]^{-1}\right\}$$

Here, you want to show that this update *minimizes* a certain norm $\|E\|$ where $E = \left[H^{(n+1)}\right]^{-1} - \left[H^{(n)}\right]^{-1}$, subject to the following two constraints:

- The secant condition: $H^{(n+1)}\delta = \gamma \implies \left[H^{(n+1)}\right]^{-1}\gamma = \delta = \left[H^{(n)}\right]^{-1}\gamma + E\gamma \implies \boxed{E\gamma = r}$, where $r = \delta - \left[H^{(n)}\right]^{-1}\gamma$.
- $H^{(n+1)}$ is still symmetric, hence $\boxed{E = E^T}$.

In particular, we consider the weighted Frobenius norm

$$F(E) = \|E\|^2 = \|MEM^T\|_F^2 = \text{tr}(ME^TM^TMEM^T) = \text{tr}(M^TMEM^TME^T) = \boxed{\text{tr}(WEWE^T)}$$

where M is some full-rank matrix and W is therefore a positive-definite symmetrix "weight" matrix *to be determined*. Strong duality holds because F is a convex quadratic function of E and the constraints are affine. Furthermore, the problem is convex and satisfies Slater's condition (there are no inequality constraints), so the KKT equations ($\partial L/\partial E = 0$ + feasibility) are necessary and sufficient conditions for optimality. Hence we can solve the KKT equations to determine E:

(a) Write down the Lagrangian in the form $L(E, \lambda, \Gamma) = \text{tr}(\cdots)$ for minimizing $F(E)$ subject to these constraints, where $\lambda \in \mathbb{R}^N$ are the Lagrange multipliers for $E\gamma = r$ and $\Gamma^T \in \mathbb{R}^{N\times N}$ are Lagrange multipliers for the N^2 constraints (albeit half redundant) $E = E^T$ [i.e. they should give a term: $\text{tr}\,\Gamma(E - E^T)$].

(b) Obtain the equation $0 = \frac{\partial L}{\partial E}$ from $0 \approx L(E+\Delta) - L(E)$ for arbitrary matrix Δ, dropping $O(\Delta^2)$ terms. Solve for E in terms of λ and Γ.

(c) From $E = E^T$, solve for $\Gamma^T - \Gamma$ and combine with your previous part to find E in terms of λ alone.

(d) From $E\gamma = r$, solve for $\lambda =$ (some expression involving $\lambda^T W^{-1}\gamma$). Hence, solve for $\lambda^T W^{-1}\gamma$ in terms of r, γ, W. Combine with the previous part to solve for E in terms of r, γ, W.

(e) If we choose $W = H^{(n+1)}$ (which is symmetric and, as shown in class, positive definite if δ came from a sufficiently good line minimization), the secant condition gives $W^{-1}\gamma = \delta$. Show that this makes your E formula yield the BFGS update for $\left[H^{(n)}\right]^{-1}$.

[Problem is based on Greenstadt and Goldfarb (1970).]