Math 560
Optimization

Homework I

Frédéric Boileau

Prof. Michael Rabbat

23rd January 2017

1

Compute the gradient and Hessian of the Rosenbrock function

$$f(x) = 100(x_2 - x_1)^2 + (1 - x_1)^2$$

$$\nabla f(x) = (2(200x_1^3 - 200x_1x_2 + x_1 - 1), 200(x_2 - x_1^2))$$

$$\nabla^2 f(x) = \begin{Bmatrix} -400(x_2 - x_1^2) + 800x_1^2 + 2 & -400x_1 \\ -400x_1 & 200 \end{Bmatrix}$$

We now consider the point $x* = (1, 1)$. We have shown that a necessary condition for a point to be a local minimizer of a function is that the gradient vanishes there. By direct substitution we can see that this is the case for x^*. Moreover

$$\nabla f(x) = 0 \iff f_2(x) = 0 \iff x_1 = \pm\sqrt{x}_2 \quad x_2 > 0 \tag{1}$$

We now show that the function is convex which means that is there is another local minimizer there must be in fact a whole neighborhood of local minimizers around x^*, which is obviously false by the continuity of the condition 1. Let us look at the characteristic polynomial of the Hessian.

$$\rho(\lambda, x_1, x_2) = (200 - \lambda)(-400(x_2 - x_1^2) + 800x_1^2 + 2 - \lambda) - 1600x_1^2 \tag{2}$$

Now from 1 we have :

$$\rho = (200 - \lambda)(800x_2 + 2 - \lambda) - 1600x_2 \quad , x_2 > 0 \tag{3}$$

We can solve this last expression for lambda. It has two roots expressed in terms of x_2. In both cases λ is positive whenever x_2 is positive. This means the eigenvalues are positive for all $x \in \mathbb{R}^2$. Hence the Hessian is positive definite and the function is thus convex. As seen in class this means there is only one minimizer (both local and global), namely x^*.

2

Let f be a convex function, show that the set of its global minimizers is convex. Let ϕ and θ be any two global minimizers. Clearly

$$f(\theta) = f(\phi) \leq f(x) \quad \forall\ x \in X \tag{4}$$

Moreover, by the definition of convexity we have that

$$f(\lambda x + \mu y) \leq \lambda f(x) + \mu f(y) \quad \forall\ \lambda, \mu \in \mathbb{R} \quad x, y \in X \tag{5}$$

Let $t \in (0, 1)$

$$\begin{aligned} f(t\phi + (1-t)\theta) &\leq tf(\phi) + (1-t)f(\theta) \\ &= f(\theta) = f(\phi) \end{aligned}$$

Hence the set of global minimizers contains all it convex combinations which is equivalent to saying it is a convex set.

3

Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable. Let $g(y) = f(Ay + b)$ for a given matrix $A \in \mathbb{R}^{n\times n}$ and vector $b \in \mathbb{R}^n$.

$$D_i\ g(y) = f\left(\sum_i e_i \sum_j a_{ij} y_j + b\right) \tag{6}$$

Define $x_i = \sum_j a_{ij} y_j + b$ and $a = Ay + b$. Then we can rewrite the expression above :

$$\frac{\partial}{\partial y_i} g(y) = \frac{\partial g(y)}{\partial x_i}\frac{\partial x_i}{\partial y_i} \tag{7}$$

$$= \frac{\partial f(\sum_i x_i + b)}{\partial x_i} \times \sum_i a_{ij} \tag{8}$$

$$= \sum_i a_{ij}\ D_i(f(y))|_a \tag{9}$$

$$\tag{10}$$

We have the result, let us rewrite it in matrix notation:

$$\nabla g(y) = \sum_j e_j \sum_i a_{ij}\ D_i(f(y))|_a$$

$$= A^\intercal \nabla f(y)|_a$$

For the Hessian we simply iterate the procedure.

$$D_i\ g(y) = D_i(A^\intercal D_i f(y)|_a) \tag{11}$$

$$= A^\intercal (D_i D_i f(y)|_a) \tag{12}$$

$$= A^\intercal (D_i (D_i f(y)|_a)^\intercal) \tag{13}$$

$$= A^\intercal (D_i (A^\intercal (D_i f(y))^\intercal|_a) \tag{14}$$

$$= A^\intercal D_i^2\ f(y)^\intercal|_a\ A \tag{15}$$

Where the partial derivative has been treated as a linear operator and the transpose was taken for conformality. In matrix notation this gives the desired result, i.e. :

$$\nabla^2 g(y) = A^\intercal \nabla^2 f(y)|_a\ A \tag{16}$$

4

Prove the following functions are convex.
First note that the second derivative is a special case of the Hessian and that $f''(x) \geq 0$ is equivalent to $\nabla^2 f(x) \succeq 0$ for proving convexity.

- Let $f(x) = e^{ax} \Rightarrow f''(x) = a^2 e^{ax} > 0 \quad \forall x \in \mathbb{R}$ and hence the function is convex for all x
- Let $f(x) = \|x\|_2$, first we note that the norm is nonnegative for any real valued vector. Moreover squaring over positive values preserves monotonicity. Hence we can prove the relaxed statement $g(x) = \|x\|_2^2$ is convex.

$$g(x) := \sum_{i=1}^{n} x_i^2 \tag{17}$$

$$\frac{\partial}{\partial x_i} g(x) = \frac{\partial}{\partial x_i} \sum_{k=1}^{n} x_k^2 = 2x_i \tag{18}$$

$$\nabla g(x) = \sum_{i=1}^{n} 2x_i e_i \quad \Rightarrow \quad H(g(x)) = 2I \succ 0 \tag{19}$$

- The sum of two convex functions is convex.

$$\begin{aligned}
f(z) &= f_1(z) + f_2(z) \\
f(\lambda x + \mu y) &= f_1(\lambda x + \mu y) + f_2(\lambda x + \mu y) \\
&\leq \lambda(f_1 + f_2)(x) + \mu(f_1 + f_2)(y) = \lambda f(x) + \mu f(y)
\end{aligned}$$

- Suping over functions preserves convexity Let $f(x) := \max\{f_1(x), f_2(x)\}$ where f_1 and f_2 are both convex functions from $\mathbb{R}^n$ to the reals.

  Let $\theta := \lambda x + (1-\lambda)x \quad \lambda \in (0,1)$

  $$\begin{aligned} f(\theta) &= \max\{f_1(\theta), f_2(\theta)\} && \text{definitions} \\ &\leq \max\{\lambda f_1(x) + (1-\lambda)f_1(x), \lambda f_2(x) + (1-\lambda)f_2(x)\} && \text{convexity of } f_1 \text{ and } f_2 \\ &\leq \lambda \max\{f_1(x), f_2(x)\} + (1-\lambda)\max\{f_1(x), f_2(x)\} && \text{generalized triangle inequality} \\ &= \lambda f(x) + (1-\lambda)f(x) \end{aligned}$$

  Another neater proof is that a function is convex if and only if its epigraph is convex. A function defined by taking the pointwise supremum as for its epigraph the intersection of the epigraphs of the functions we're supping over. Hence $\text{epi}f = \text{epi}f_1 \cap \text{epi}f_2$. Moreover intersection preserves convexity. So $\text{epi}f$ is convex and hence f is.

- Composition with affine mapping preserves convexity

  Let $h : \mathbb{R}^m \to \mathbb{R}^n$ be defined the following way:

  $$y \mapsto Ay + b \qquad A \in \mathbb{R}^{n \times m},\ b \in \mathbb{R}^m$$

  h is an affine map. Let $f : \mathbb{R}^n \to \mathbb{R}$ be any convex function and define $g := f \circ h$. We show that g is a convex function.

  $$\begin{aligned} g(\lambda x + (1-\lambda)y) :&= (f \circ h)(\lambda x + (1-\lambda)y) && \lambda \in (0,1) \\ &= f(h(\lambda x + (1-\lambda)y)) \\ &= f(\lambda h(x) + (1-\lambda)h(y)) && \text{characteristic property of affine maps} \\ &\leq \lambda f(h(x)) + (1-\lambda)f(g(y)) && \text{which is what we wanted to show} \end{aligned}$$