Fenchel Conjugates and Newton Steps

R. Kyng & M. Probst Gutenberg *Problem Set 9 — Wednesday, May 10th*

The exercises for this week will not count toward your grade, but you are highly encouraged to solve them all. This exercise sheet has exercises related to week 9. We encourage you to start early so you have time to go through everything.

To get feedback, you must hand in your solutions by 23:59 on May 18. Both hand-written and LaTeX solutions are acceptable, but we will only attempt to read legible text.

Exercise 1: Fenchel Conjugate Examples

- Consider the function $\mathcal{E}(y) = y^4/4$, where $\mathcal{E} : \mathbb{R} \to \mathbb{R}$. Compute the Fenchel conjugate $\mathcal{E}^*$ of $\mathcal{E}$. Verify the gradient and Hessian relationships between these that we discussed in class.
- Consider the function $\mathcal{E}(\boldsymbol{y}) = \frac{1}{2}\left\|\boldsymbol{y}\right\|^2$, where $\|\cdot\|$ is an arbitrary norm. Compute the Fenchel conjugate $\mathcal{E}^*$ the function.
  Hint: express your answer using the dual norm
  $$\left\|\boldsymbol{z}\right\|_* = \max_{\substack{\boldsymbol{y} \text{ s.t.} \\ \|\boldsymbol{y}\| \leq 1}} \boldsymbol{z}^\top \boldsymbol{y}.$$
- Consider the function $\mathcal{E}(\boldsymbol{y}) = \|\boldsymbol{y}\|$, where $\|\cdot\|$ is an arbitrary norm. Compute the Fenchel conjugate $\mathcal{E}^*$ the function. *Warning: you may get ∞ somewhere....*
- *UPDATED.* Consider the function $\mathcal{E} : \mathbb{R} \to \mathbb{R}$ given by
  $$\mathcal{E}(y) = \begin{cases} y^{2.2}/(2.2) \text{ if } |y| \leq 1 \\ |y|^{1.1} - (1 - 1/2.2)) \text{ if } |y| > 1 \end{cases}$$
  Compute its Fenchel conjugate $\mathcal{E}^*$.
- *BONUS (NEW).* Consider the function $\mathcal{E} : \mathbb{R} \to \mathbb{R}$ given by
  $$\mathcal{E}(y) = \begin{cases} y^{2.2}/(2.2) \text{ if } |y| \leq 1 \\ |y|^{1.1}/1.1 - (1/1.1 - 1/2.2)) \text{ if } |y| > 1 \end{cases}$$
  Compute its Fenchel conjugate $\mathcal{E}^*$.

Exercise 2: Fenchel Twice

- Consider a convex function $\mathcal{E} : \mathbb{R}^n \to \mathbb{R}$ with Fenchel conjugate $\mathcal{E}^*$. Suppose $\boldsymbol{\nabla}\mathcal{E}$ is a bijection. Explain why $\boldsymbol{\nabla}\mathcal{E}^*$ is the inverse function of $\boldsymbol{\nabla}\mathcal{E}$.
- Sketch a proof that given a convex function $\mathcal{E} : \mathbb{R}^n \to \mathbb{R}$ with Fenchel conjugate $\mathcal{E}^*$, we have $(\mathcal{E}^*)^* = \mathcal{E}$. I.e. the Fenchel conjugate transformation applied twice returns the original function. You may impose any technical conditions on $\mathcal{E}$ that you like.

Exercise 3: Fenchel Transformation

Consider a convex function $\mathcal{E} : \mathbb{R}^n \to \mathbb{R}$ with Fenchel conjugate $\mathcal{E}^*$.

- Give a convenient expression for the Fenchel conjugate of $\widehat{\mathcal{E}}(\boldsymbol{y}) = \mathcal{E}(\boldsymbol{y} + \boldsymbol{t})$, where $\boldsymbol{t}$ is a fixed vector.
- Give a convenient expression for the Fenchel conjugate of $\widehat{\mathcal{E}}(\boldsymbol{y}) = \mathcal{E}(\boldsymbol{M}\boldsymbol{y})$, where $\boldsymbol{M}$ is an invertible matrix.
- Give a convenient expression for the Fenchel conjugate of $\widehat{\mathcal{E}}(\boldsymbol{y}) = \mathcal{E}(\boldsymbol{y}) + \boldsymbol{g}^\top \boldsymbol{y}$, where $\boldsymbol{g}$ is a fixed vector.

Exercise 4: Newton Steps are affine invariant

Consider a convex function $\mathcal{E}(\boldsymbol{y}) : \mathbb{R}^n \to \mathbb{R}$.

Suppose that starting from $\boldsymbol{y}_0$ we take a Newton step given by $\boldsymbol{\delta} = -H_{\mathcal{E}}(\boldsymbol{y}_0)^{-1} \boldsymbol{\nabla}\mathcal{E}(\boldsymbol{y}_0)$.

Next, consider the reparameterized function $\widehat{\mathcal{E}}(\boldsymbol{x}) = \mathcal{E}(\boldsymbol{M}\boldsymbol{x})$, where $\boldsymbol{M}$ is invertible. Suppose that starting from $\boldsymbol{x}_0 = \boldsymbol{M}^{-1}\boldsymbol{y}_0$ we take a Newton step given by $\widehat{\boldsymbol{\delta}} = -H_{\widehat{\mathcal{E}}}(\boldsymbol{x}_0)^{-1} \boldsymbol{\nabla}\widehat{\mathcal{E}}(\boldsymbol{x}_0)$.

- Show that $\boldsymbol{y}_0 + \boldsymbol{\delta} = \boldsymbol{M}(\boldsymbol{x}_0 + \widehat{\boldsymbol{\delta}})$.
- Suppose our updates $\boldsymbol{\delta}$ came from gradient descent instead of Newton steps – would a similar relationship still hold? What if $\boldsymbol{M}$ is orthonormal?