Assignment 3

Convex Optimization SS 2018

Patrick Knöbelreiter, knoebelreiter@icg.tugraz.at

23.05.2018

Submission: Upload your solution as a PDF and your implementation as PY files to the TeachCenter. You can also use your scanned hand-written manuscript.

Deadline: June 11^{th}, 2018 at 23:55h.

Infimal Convolution

Consider the inf-convolution between two convex functions f, g on $\mathbb{R}^n$

$$\phi(x) = \inf_y f(y) + g(x - y) \tag{1}$$

If $g(x-y)$ has the special form $g(x-y) = \frac{1}{2\lambda}\|x-y\|^2$, the function $\phi(x)$ is called *Moreau Envelope* of f with parameter $\lambda > 0$. The Moreau envelope can be used to obtain a smooth approximation of non-smooth functions. Compute the Moreau envelope of the following exercises and plot the original function $f(y)$ as well as the envelope $\phi(x)$ for different values of λ.

1. $f(y) = \max(0, 1-y), \quad y \in \mathbb{R}$
2. $f(y) = |y| + {}^1\!/_3|y|^3, \quad y \in \mathbb{R}$
3. $f(y) = \begin{cases} 0 & \text{if } y \in [a,b] \\ \infty & \text{else} \end{cases} \qquad a, b \in \mathbb{R}, \quad y \in \mathbb{R}$
4. $f(y) = \max(|y|, y^2), \quad y \in \mathbb{R}$

Proximal Operators

Recall the definition of the proximal operator:

$$\text{prox}_{\lambda f}(x) = \arg\min_u \left(f(u) + \frac{1}{2\lambda}\|u - x\|_2^2 \right), \quad \lambda > 0.$$

1. Calculate the proximal operator for the following functions. For the exercises 1(d),(e),(f) additionally plot $f(x)$ and $\text{prox}_{\lambda f}(x)$ with $\lambda = 1$.
   a) Max-norm: $f(x) = \|x\|_\infty, \quad x \in \mathbb{R}^n$ Hint: Use the Moreau identity.
   b) Zero-norm: $f(x) = \|x\|_0, \quad x \in \mathbb{R}^n$
   c) Elastic net: $f(x) = \|x\|_1 + \frac{a}{2}\|x\|_2^2,\ a > 0, \quad x \in \mathbb{R}^n$
   d) $f(x) = \begin{cases} -a\ln(x) + \frac{x^2}{2} & \text{for } x > 0 \\ +\infty & \text{for } x \le 0 \end{cases} \quad a > 0, \quad x \in \mathbb{R}$
   e) $f(x) = \max(|x| - a, 0), \quad a > 0, \quad x \in \mathbb{R}$
   f) $f(x) = \begin{cases} \ln(a) - \ln(a - |x|) & \text{for } |x| < a \\ +\infty & \text{otherwise} \end{cases} \quad a > 0, \quad x > \mathbb{R}$
2. Prove the following version of Moreau's identity

$$x = \text{prox}_{\lambda f^*}(x) + \lambda\, \text{prox}_{f/\lambda}({}^x\!/_\lambda)$$

(F)ISTA algorithm for LASSO

We want to solve the following instance of a least absolute shrinkage and selection operator (LASSO) model:

$$\min_{C} E(C) := \frac{1}{2}||DC - B||_2^2 + \lambda||C||_1, \tag{2}$$

where $\{D \in \mathbb{R}^{MN \times L} : ||D_l||_2 \leq 1,\ l = 1, \ldots, L\}$ is a known dictionary, $C \in \mathbb{R}^{L \times K}$ is the coefficient matrix we want to determine and $B \in \mathbb{R}^{MN \times K}$ is the training data.

In our case each column of B is a vectorized training image of size $M \times N$ from the MNIST database[1], the columns of D are the dictionary atoms and each column of C contains the sparse code for one of the images w.r.t. the dictionary D. In total there are K images and typically we have $L < MN$. Hence, we have an instance of a sparse coding problem, where the compression comes from the fact that using the optimal basis D the data has a lower-dimensional representation.

For solving the LASSO we are going to use the iterative shrinkage-thresholding algorithm (ISTA) and the accelerated version called Fast ISTA (FISTA).

The (F)ISTA algorithm can be used to optimize problems of the following form:

$$\min_{x} f(x) + g(x), \quad x \in \mathbb{R}^n \tag{3}$$

where $g(x)$ is a convex, possibly non-smooth function and $f(x)$ is a smooth convex function.

ISTA

The iterates of the ISTA algorithm for $k \geq 1$ and a Lipschitz constant L of ∇f are given by

$$x^{k+1} = \text{prox}_{g/L}(x^k - \tfrac{1}{L}\nabla f(x^k)) \tag{4}$$

FISTA

The algorithm computes a minimizer by iterating for $k \geq 1$, starting values $y^1 = x^0 \in \mathbb{R}^n$, $t^1 = 1$ and L a Lipschitz constant of ∇f

$$\begin{cases} x^k = \text{prox}_{g/L}(y^k - \frac{1}{L}\nabla f(y^k)) \\ t^{k+1} = \frac{1+\sqrt{1+4(t^k)^2}}{2} \\ y^{k+1} = x^k + \frac{t^k - 1}{t^{k+1}}(x^k - x^{k-1}) \end{cases} \tag{5}$$

Tasks

1. Compute the Lipschitz constant used in (4) and (5). You are allowed to use all `numpy` functions for this task.
2. Compute the proximal map used in (4) analytically.
3. Implement the ISTA algorithm
4. Implemnet the FISTA algorithm
5. For both algorithms
   a) Use your implementation of (F)ISTA to solve problem (2) with the pre-computed dictionaries.
   b) Plot the energy $E(C)$ over the iterates and compare the energies in one plot.
   c) Reconstruct 5 samples using your optimized coefficients and plot them along with the original samples.

Framework

The dictionary D is already pre-computed and can be directly used. There exist 4 different versions with a different L. For example the file `D64.npy` is the dictionary D with $L = 64$. You can use the following code-snippet to load the provided data:

```
B = np.load('mnist_training_images.npy') / 255.0
D = np.load('D64.npy')
```

[1] http://yann.lecun.com/exdb/mnist/