Convex Optimization for Computer Vision
Lecture: M. Möller and T. Möllenhoff
Exercises: E. Laude
Summer Semester 2016

Computer Vision Group
Institut für Informatik
Technische Universität München

Weekly Exercises 6

Room 01.09.014
Friday, 3.6.2016, 09:00-11:00
Submission deadline: Wednesday, 1.6.2016, 14:00, Room: 02.09.023

Theory: Line Search (6 Points)

Exercise 1 (6 Points)**.** In the lecture, we have seen that for $E : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$

$$E(u) = G(u) + F(u),$$

with closed, proper, convex $G : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ and L-smooth $F : \mathbb{R}^n \to \mathbb{R}$ the proximal gradient algorithm given by the iteration

$$u^{k+1} = \mathrm{prox}_{\tau G}(u^k - \tau \nabla F(u^k)),$$

converges with rate $\mathcal{O}(1/k)$, i.e.

$$E(u^k) - E(u^*) \in \mathcal{O}(1/k).$$

In this exercise, we select τ_k using a line search (In practice this is helpful if L is not known): The line search works as follows: Start at some initial $\tau := \hat{\tau} > 0$ and iteratively compute $\tau := \beta\tau$ for $0 < \beta < 1$ until the following inequality, that we have seen in the lecture holds:

$$F(u - \tau\varphi_\tau(u)) \le F(u) - \tau\langle \nabla F(u), \varphi_\tau(u)\rangle + \frac{\tau}{2}\left\|\varphi_\tau(u)\right\|^2.$$

Prove that proximal gradient with line search converges with rate $\mathcal{O}(1/k)$.
Hint: Show that the step size τ selected by the line search satisfies

$$\tau \ge \tau_{\min} = \min\{\hat{\tau}, \beta/L\}.$$

Solution. To formalize the line search algorithm, we can state that

$$\begin{aligned} \tau^k =& \beta^{j_k}\hat{\tau} \quad \text{ for } j_k = \min_j \{j \in \mathbb{B} \mid H(u^k, \beta^j\hat{\tau}) \le 0\} \\ H(u,\tau) :=& F(u - \tau\phi_\tau(u)) - F(u) + \tau\langle \nabla F(u), \phi_\tau(u)\rangle - \frac{\tau}{2}\|\phi_\tau(u)\|^2 \end{aligned}$$

Part 1: Show that $\tau^k \ge \tau_{min} = \min(\hat{\tau}, \frac{\beta}{L})$ for all k.

- Let $\hat{\tau} \le \frac{1}{L}$. It has been shown in the lecture, that $H(u,\tau) \le 0$ hold for all u if $\tau \le \frac{1}{L}$. Thus, $j_k = 0$ for all k, and $\tau_k = \hat{\tau}$, which means $\tau_k \ge \hat{\tau}$.

- Now let $\hat{\tau} > \frac{1}{L}$. Assume there exists a $k \geq 1$ with $\tau_k < \frac{\beta}{L}$. Then

$$\frac{\beta^{j_k}}{\hat{\tau}} < \frac{\beta}{L},$$

  which means

$$\beta^{j_k - 1}\hat{\tau} < \frac{1}{L}.$$

  Again, according to the lecture we must have $H(u^k, \beta^{j_k-1}\hat{\tau}) \leq 0$, which contradicts the definition of j as the smallest natural number for which the previous inequality holds.

Part 2: Show that $E(u^N) - E(u^*) \in \mathcal{O}(1/k)$:

It was shown in the lecture that $H(u^k, \tau^k) \leq 0$ implies

$$E(u^{k+1}) = E(u^k - \tau^k \phi_{\tau^k}(u)) \leq E(w) + \langle \phi_{\tau^k}(u^k), u^k - w\rangle - \frac{\tau^k}{2}\|\phi_{\tau^k}(u^k)\|^2$$

for arbitrary elements w.

By inserting $w = u^k$, one quickly sees that the energy is monotonically decreasing.

By inserting a minimizer u^* of the energy E, i.e. $w = u^*$, and completing the square similar to the lecture, we find

$$E(u^{k+1}) - E(u^*) \leq \frac{1}{2\tau^k}\left(\|u^k - u^*\|^2 - \|u^{k+1} - u^*\|^2\right).$$

We sum the above inequality from $k = 0$ to $k = N - 1$ and obtain

$$\begin{aligned}
N(E(N) - E(u^*)) \leq & \sum_{k=0}^{N-1}(E(u^{k+1}) - E(u^*)) \\
\leq & \sum_{k=0}^{N-1} \frac{1}{2\tau^k}\left(\|u^k - u^*\|^2 - \|u^{k+1} - u^*\|^2\right) \\
\leq & \frac{1}{2\tau_{min}} \sum_{k=0}^{N-1}\left(\|u^k - u^*\|^2 - \|u^{k+1} - u^*\|^2\right) \\
= & \frac{1}{2\tau_{min}}\left(\|u^0 - u^*\|^2 - \|u^N - u^*\|^2\right) \\
\leq & \frac{1}{2\tau_{min}}\|u^0 - u^*\|^2,
\end{aligned}$$

where we used the monotonicity of the energy in the first inequality. After dividing by N we see that the energy is approaching the minimum with a rate of at least $1/N$.

Programming: TV Deblurring (12 Points)

Exercise 2 (12 Points)**.** Given a blurry and noisy input image f, reconstruct a sharper image u^* by solving the following optimization problem

$$u^* = \arg\min_u \frac{1}{2} \|k * u - f\|^2 + \alpha \|Du\|_{2,1} \tag{1}$$

with proximal gradient descent (for the definition of the convolution $k * u$, see the lecture slides). To do so, perform the following steps:

- First construct a convolution kernel k of your choice, for example by using the MATLAB command `fspecial`.
- Then build a sparse matrix representing the convolution with an image u. You can use any boundary condition, and feel free to use `convmtx2`.
- Given the image `flowers.png`, construct a blurred and noisy version by applying your sparse matrix to it and add some Gaussian noise using `randn`.
- Restore the original image by solving (1) using proximal gradient descent. Solve the inner TV denoising problem using projected gradient descent on the dual problem.
- Experiment with different amounts of inner projected gradient descent iterations (use 15 as a starting point) and use the solution of the previous outer iteration to warm-start the algorithm.