Convex Optimization and Modeling	Jun.-Prof. Matthias Hein
Solution of Exercise Sheet 12	07.07.2010

Exercise 24 - Proximity Operator

Prove the two statements in the lecture about the proximity operator prox_f defined as

$$\operatorname{prox}_f x = \arg\min_{y \in \mathbb{R}^n} \; f(y) + \frac{1}{2}\|x - y\|^2 .$$

a. **(2 Points)** $x^* = \operatorname{prox}_f x^*$ if and only if x^* is a minimizer of f. This means the set of fixed points of prox_f is equal to the set of minimizers of f.

b. **(3 Points)** The proximity operator is non-expansive,

$$\left\|\operatorname{prox}_f x - \operatorname{prox}_f y\right\| \leq \|x - y\| .$$

Hint:

- In both cases use the optimality condition of the minimization problem. For the second problem use the same steps as for the projection onto convex sets.

Solution:

a. Suppose y^* is a minimizer of f. Then for all $y \in \mathbb{R}^n$,

$$f(y^*) = f(y^*) + \frac{1}{2}\|y^* - y^*\|^2 \leq f(y) + \frac{1}{2}\|y^* - y\| ,$$

and thus $\operatorname{prox}_f y^* = y^*$ if y^* is a minimizer of f.
The optimality condition reads,

$$0 \in \partial f(\operatorname{prox}_f x) + \operatorname{prox}_f x - x.$$

Thus if $\operatorname{prox}_f x = x$, then $0 \in \partial f(\operatorname{prox}_f x)$ and thus $\operatorname{prox}_f x$ is a minimizer of f.

b. From the optimality condition we get,

$$x - \operatorname{prox}_f x \in \partial f(\operatorname{prox}_f x).$$

Using the first-order condition we get,

$$\begin{aligned} f(\operatorname{prox}_f y) &\geq f(\operatorname{prox}_f x) + \left\langle \partial f(\operatorname{prox}_f x), \operatorname{prox}_f y - \operatorname{prox}_f x \right\rangle \\ &= f(\operatorname{prox}_f x) + \left\langle x - \operatorname{prox}_f x, \operatorname{prox}_f y - \operatorname{prox}_f x \right\rangle . \end{aligned}$$

In the same way we get,

$$f(\operatorname{prox}_f x) \geq f(\operatorname{prox}_f y) + \left\langle y - \operatorname{prox}_f y, \operatorname{prox}_f x - \operatorname{prox}_f y \right\rangle .$$

Adding both equations from the other one we get,

$$0 \geq \left\langle x - y + \operatorname{prox}_f y - \operatorname{prox}_f x, \operatorname{prox}_f y - \operatorname{prox}_f x \right\rangle = \left\langle x - y, \operatorname{prox}_f y - \operatorname{prox}_f x \right\rangle + \left\|\operatorname{prox}_f y - \operatorname{prox}_f x\right\|^2 ,$$

In total, we get

$$\left\|\operatorname{prox}_f y - \operatorname{prox}_f x\right\|^2 \leq \left\langle x - y, \operatorname{prox}_f y - \operatorname{prox}_f x \right\rangle \leq \|x - y\| \left\|\operatorname{prox}_f y - \operatorname{prox}_f x\right\| ,$$

which yields the desired result.

One can also derive a stronger result as follows,

$$
\begin{aligned}
0 &\geq 2\left\langle x-y, \operatorname{prox}_f y - \operatorname{prox}_f x\right\rangle + 2\left\|\operatorname{prox}_f y - \operatorname{prox}_f x\right\|^2 \\
\|x-y\|^2 &\geq \|x-y\|^2 + 2\left\langle x-y, \operatorname{prox}_f y - \operatorname{prox}_f x\right\rangle + 2\left\|\operatorname{prox}_f y - \operatorname{prox}_f x\right\|^2 \\
\|x-y\|^2 &\geq \left\|(x-\operatorname{prox}_f x) - (y - \operatorname{prox}_f y)\right\|^2 + \left\|\operatorname{prox}_f y - \operatorname{prox}_f x\right\|^2
\end{aligned}
$$

Exercise 25 - Total Variation for general graphs

The goal of this exercise is to implement the total variation denoising scheme for general graphs developed in the lecture.

a. **(3 Points)** Prove that $\|A\|^2 \leq 4\max_r \sum_{j=1}^n w_{rj}^2$.

b. **(1 Point)** Show that $(A^T f)_{ij} = w_{ij}(f_i - f_j)$.

c. **(7 Points)** Implement the total variation method as introduced in the lecture as

```
f=TotalVariation(MAXITER,epsT,Y,W,lambda)
```

where $Y, f \in \mathbb{R}^{nm}$ are the input and output images represented as a vectors (use `reshape`), $W \in \mathbb{R}^{mn\times mn}$ encodes the weights between the pixels, and `lambda` is the regularization parameter. Use the box constraints $C = [0,1]^n$.

- use $(f(x^k) - f(x^{k+1}))/f(x^k) < 10^{-5} =$ `epsT` as stopping criterion or number of steps larger than `MAXITER` $= 1000$,
- Analyze the dependence of the number of required iterations on the employed regularization parameter λ.
- Which λ yields the best results in terms of reconstruction error and which in terms of visual appearance ?

Hints:

- make sure that all the matrices you use have sparse format (generate them with the command `sparse`). You can obtain the row and column indices, and the values of the non-zero entries of a sparse matrix via: `[ix,jx,val]=sparse(W)`.

Send the matlab-code and all plots (as `png`-files) to Shyam Rangapuram, email: `r.shyamsundar@gmail.com`.

Solution:

- The proof mainly uses Cauchy-Schwarz together with upper-bounding a sum of positive terms

$$
\begin{aligned}
\|A\alpha\|_2^2 = 4\sum_{i=1}^n \Big(\sum_{j=1}^n w_{ij}\alpha_{ij}\Big)^2 &\leq 4\sum_{i=1}^n \sum_{j=1}^n w_{ij}^2 \sum_{j=1}^n \alpha_{ij}^2 \\
&\leq 4\max_{r=1,\ldots,n} w_{rj}^2 \sum_{i,j=1}^n \alpha_{ij}^2 = 4\max_{r=1,\ldots,n} w_{rj}^2 \, \|\alpha\|_2^2 .
\end{aligned}
$$

Thus with $\|A\|_{2,2}^2 = \sup_\alpha \frac{\|A\alpha\|_2^2}{\|\alpha\|_2^2} \leq 4\max_{r=1,\ldots,n} w_{rj}^2$.

- The definition of the transpose (or more general the adjoint) operator is

$$\langle A\alpha, f\rangle_{\mathbb{R}^V} = \left\langle \alpha, A^T f\right\rangle_{\mathbb{R}^E}, \quad \forall \alpha \in \mathbb{R}^E,\ f \in \mathbb{R}^V.$$

  Thus,

$$\begin{aligned}
\langle A\alpha, f\rangle_{\mathbb{R}^V} &= 2\sum_{i,j=1}^{n} w_{ij}\alpha_{ij} f_i = \sum_{i,j=1}^{n} w_{ij}\alpha_{ij} f_i - \sum_{i,j=1}^{n} w_{ij}\alpha_{ji} f_i \\
&= \sum_{i,j=1}^{n} w_{ij}\alpha_{ij} f_i - \sum_{j,i=1}^{n} w_{ji}\alpha_{ij} f_j \\
&= \sum_{i,j=1}^{n} \alpha_{ij}\, w_{ij}(f_i - f_j).
\end{aligned}$$

  Thus $(A^T f) = w_{ij}(f_i - f_j)$.

- The number of required iterations increases with increasing λ. This has a theoretical backup as the Lipschitz-constant of Ψ is

$$L = 4\lambda^2 \max_{r=1,\dots,n} w_{rj}^2.$$

  Thus the Lipschitz constant increases quadratically and we have the bound,

$$f(x^{(k)}) - p^* \leq \frac{2\,L}{(k+1)^2} \left\| x^{(0)} - x^* \right\|^2,$$

  where the approximation guarantee decreases with increasing L and thus it can be expected that one stops earlier (even though the stopping criterion is not directly linked to this condition).