Convex Optimization for Machine Learning and Computer Vision

Lecture: Dr. Tao Wu
Exercises: Yuesong Shen, Zhenzhang Ye
Winter Semester 2018/19

Computer Vision Group
Institut für Informatik
Technische Universität München

Weekly Exercises 10

Room: 01.09.014
Wednesday, 16.01.2019, 12:15-14:00
Submission deadline: Monday, 14.01.2019, 16:15, Room 01.09.014

Theory: Proximal algorithms (8+4 Points)

Exercise 1 (6 points)**.** (Variant of Douglas-Rachford splitting) Let $F, G : \mathbb{E} \to \bar{\mathbb{R}}$ be 2 convex, proper and lower semi continuous (l.s.c.) functions. We consider the following updates for solving $\min_u F(u) + G(u)$:

$$\begin{cases} u^{k+1} = \operatorname{argmin}_u G(u) + \frac{\epsilon}{2}\|u - u^k\|^2 + \frac{1}{2\tau}\|u - v^k\|^2, \\ v^{k+1} = v^k - u^{k+1} + \operatorname{prox}_{\tau F}(2u^{k+1} - v^k). \end{cases} \tag{1}$$

Reformulate Eq. 1 as a customized proximal iteration.

Solution. Denote $p^k := (u^k - v^k)/\tau$. We have that:

$$\begin{aligned} u^{k+1} &= \operatorname{argmin}_u G(u) + \frac{\epsilon}{2}\|u - u^k\|^2 + \frac{1}{2\tau}\|u - v^k\|^2 && (2) \\ &= \operatorname{argmin}_u G(u) + \frac{1+\epsilon\tau}{2\tau}\|u - \frac{v^k + \epsilon\tau u^k}{1+\epsilon\tau}\|^2 && (3) \\ &= \operatorname{argmin}_u G(u) + \frac{1+\epsilon\tau}{2\tau}\|u - (u^k - \frac{\tau}{1+\epsilon\tau}p^k)\|^2 && (4) \\ &= \operatorname{prox}_{\tau/1+\epsilon\tau G}(u^k - \frac{\tau}{1+\epsilon\tau}p^k). && (5) \end{aligned}$$

Thus

$$\begin{aligned} u^k - \frac{\tau}{1+\epsilon\tau}p^k &\in (I + \frac{\tau}{1+\epsilon\tau}\partial G)(u^{k+1}), && (6) \\ 0 &\in (\frac{1}{\tau} + \epsilon)(u^{k+1} - u^k) + p^k + \partial G(u^{k+1}). && (7) \end{aligned}$$

From the slides we have that

$$\begin{aligned} v^{k+1} &= v^k - u^{k+1} + \operatorname{prox}_{\tau F}(2u^{k+1} - v^k) && (8) \\ \Longleftrightarrow 0 &\in \tau(p^{k+1} - p^k) + \partial F^*(p^{k+1}) - (2u^{k+1} - u^k). && (9) \end{aligned}$$

Therefore Eq. 1 can be reformulated as the following customized proximal iteration:

$$0 \in \begin{bmatrix} (\frac{1}{\tau} + \epsilon)I & -I \\ -I & \tau I \end{bmatrix} \begin{bmatrix} u^{k+1} - u^k \\ p^{k+1} - p^k \end{bmatrix} + \begin{bmatrix} \partial G & I \\ -I & \partial F^* \end{bmatrix} \begin{bmatrix} u^{k+1} \\ p^{k+1} \end{bmatrix} \tag{10}$$

Exercise 2 (6 points)**.** (PDHG derivation for optical flow estimation) Consider the optimization problem given by Eq. 38 in this week's programming assignment: it has the form $f(U) + g(KU)$ where

$$f : U \mapsto \sum_{i=1}^{mn} \lambda |\langle (D\tilde{I}_1)_i, (U - \bar{U})_i \rangle + (\tilde{I}_1 - I_0)_i|, \tag{11}$$

$$g : V \mapsto \|V\|_{1,2} = \sum_{i=1}^{2mn} \|V_i\|_2, \tag{12}$$

$$K = \nabla. \tag{13}$$

Write down the PDHG updates for this problem.

Solution.

The convex conjugate of g is

$$g^* : P \mapsto \delta_C(P) \tag{14}$$

where $C = \{P \in \mathbb{R}^{2mn\times 2} | \forall 1 \leq i \leq 2mn, \|P_i\|_2 \leq 1\}$

Thus Eq. 38 can be written as the following saddle-point problem:

$$\min_U \sup_P f(U) + \langle P, KU \rangle - g^*(P) \tag{15}$$

with $\langle A, B \rangle = tr(AB^\top)$. PDHG then performs the following updates:

$$U^{k+1} = \text{argmin}_U\, f(U) + \langle K^\top P^k, U \rangle + \frac{s}{2}\|U - U^k\|_F^2 \tag{16}$$

$$P^{K+1} = \text{argmin}_P\, -\langle P, K(2U^{k+1} - U^k) \rangle + g^*(P) + \frac{t}{2}\|P - P^k\|_F^2 \tag{17}$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

Eq. 17 gives the projected gradient descent:

$$P^{K+1} = proj_C\big(P^k + \frac{1}{t}K(2U^{k+1} - U^k)\big) \tag{18}$$

with the projection being done row-wise onto unit ball (w.r.t. l_2 norm) in $\mathbb{R}^2$.

Eq. 16 can be decomposed row-wise such that for any $1 \leq i \leq mn$ we have

$$U_i^{k+1} = \text{argmin}_{U_i}\, \lambda |\langle (D\tilde{I}_1)_i, (U_i - \bar{U}_i) \rangle + (\tilde{I}_1 - I_0)_i| + \langle (K^\top P^k)_i, U_i \rangle + \frac{s}{2}\|U_i - U_i^k\|_2^2. \tag{19}$$

To lighten the notation, we define the following constants:

$$\alpha_i = (D\tilde{I}_1)_i \in \mathbb{R}^2 \tag{20}$$

$$\beta_i = (K^\top P^k)_i \in \mathbb{R}^2 \tag{21}$$

$$c_i = (\tilde{I}_1 - I_0)_i - \langle (D\tilde{I}_1)_i, \bar{U}_i \rangle \in \mathbb{R} \tag{22}$$

so that

$$U_i^{k+1} = \mathrm{argmin}_{U_i}\, \lambda|\langle \alpha_i, U_i\rangle + c_i| + \langle \beta_i, U_i\rangle + \frac{s}{2}\|U_i - U_i^k\|_2^2. \tag{23}$$

which results in the following optimality condition:

$$0 \in \lambda\partial|\cdot|(\langle \alpha_i, U_i^{k+1}\rangle + c_i)\cdot\alpha_i + \beta_i + s(U_i^{k+1} - U_i^k). \tag{24}$$

Here $\partial|\cdot|(\langle \alpha_i, U_i^{k+1}\rangle + c_i)$ denote the sub-differential of the absolute value function at point $\langle \alpha_i, U_i^{k+1}\rangle + c_i \in \mathbb{R}$. We have thus

$$\partial|\cdot|(\langle \alpha_i, U_i^{k+1}\rangle + c_i) = \begin{cases} 1 & \text{if } \langle \alpha_i, U_i^{k+1}\rangle + c_i > 0 \\ [-1,1]_i & \text{if } \langle \alpha_i, U_i^{k+1}\rangle + c_i = 0 \\ -1 & \text{if } \langle \alpha_i, U_i^{k+1}\rangle + c_i < 0 \end{cases} \tag{25}$$

We consider the 3 cases:

- when $\langle \alpha_i, U_i^{k+1}\rangle + c_i > 0$, we have

  $$0 = \lambda\alpha_i + \beta_i + s(U_i^{k+1} - U_i^k). \tag{26}$$

  thus

  $$U_i^{k+1} = U_i^k - \frac{1}{s}\beta_i - \frac{\lambda}{s}\alpha_i. \tag{27}$$

  in which case

  $$\langle \alpha_i, U_i^k - \frac{1}{s}\beta_i\rangle + c_i > \frac{\lambda}{s}\|\alpha_i\|_2^2 \tag{28}$$

- when $\langle \alpha_i, U_i^{k+1}\rangle + c_i < 0$, we have

  $$0 = -\lambda\alpha_i + \beta_i + s(U_i^{k+1} - U_i^k). \tag{29}$$

  thus

  $$U_i^{k+1} = U_i^k - \frac{1}{s}\beta_i + \frac{\lambda}{s}\alpha_i. \tag{30}$$

  in which case

  $$\langle \alpha_i, U_i^k - \frac{1}{s}\beta_i\rangle + c_i < -\frac{\lambda}{s}\|\alpha_i\|_2^2 \tag{31}$$

- when $\langle \alpha_i, U_i^{k+1}\rangle + c_i = 0$, we have

  $$U_i^{k+1} \in \{U_i^k - \frac{1}{s}\beta_i + \mu\frac{\lambda}{s}\alpha_i \mid \mu \in [-1,1]\}. \tag{32}$$

  in which case

  $$\langle \alpha_i, U_i^k - \frac{1}{s}\beta_i\rangle + c_i = -\mu\frac{\lambda}{s}\|\alpha_i\|_2^2 \tag{33}$$

Based on the case-by-case derivation above, if we denote

$$\rho_i(U) = \begin{cases} \left(\langle \alpha_i, U - \frac{1}{s}\beta_i \rangle + c_i\right) / \left(\frac{\lambda}{s}\|\alpha_i\|_2^2\right) & \text{if } \alpha_i \neq 0 \\ 0 & \text{if } \alpha_i = 0 \end{cases} \tag{34}$$

we can get

$$U_i^{k+1} = U_i^k - \frac{1}{s}\beta_i - \frac{\lambda}{s} \cdot \begin{cases} \alpha_i & \text{if } \rho_i(U_i^k) > 1 \\ \rho_i(U_i^k)\alpha_i & \text{if } |\rho_i(U_i^k)| < 1 \\ -\alpha_i & \text{if } \rho_i(U_i^k) < -1 \end{cases} \tag{35}$$

which leads to a row-wise thresholding during update.

Programming: Optical flow estimation (Due date: 28.01.2019) (12 Points)

Exercise 3 (12 Points)**.** In this exercise, we ask you to write a program in MATLAB (or Python) that solves the following inner problem for the optical flow estimation:

$$\min_{U \in \mathbb{R}^{mn \times 2}} \lambda \sum_{i=1}^{mn} |\langle (D\tilde{I}_1)_i, (U - \bar{U})_i \rangle + (\tilde{I}_1 - I_0)_i| + \|\nabla U\|_{1,2} \tag{36}$$

where $\bar{U} = (\bar{U}_x \ \bar{U}_y)$ denotes the initial estimate of the displacement field (along x and y directions), I_0 the reference image and $\tilde{I}_1$ the warped image of I_1 (which is the the image after motion) according to the displacement field $\bar{U}$, $D\tilde{I}_1 = (D_x\tilde{I}_1 \ D_y\tilde{I}_1)$, $\nabla = (D_x^\top \ D_y^\top)^\top$, The suffix i denote the $i-$th row, and $\|\cdot\|_{1,2}$ denotes the matrix $l_{1,2}$ norm (i.e. summation of l_2-norm of each row). Our aim is to find a displacement field $U = (U_x \ U_y)$ that minimizes the objective function (Eq. 38) composed of a linearized data fidelity term and a total-variation regularization term.
You are asked to solve this problem using PDHG.

Hint 1: check out exercise 0 to see how to construct D_x, D_y, ∇ etc.

Hint 2: in our case $\|\nabla\|_{spec}^2 = 8$ (optionally you can verify this by hand or numerically using `normest` in MATLAB or `scipy.sparse.linalg.svds` in Python), so you could choose e.g. $s = t = 3$ so that $s \times t \geq \|\nabla\|_{spec}^2$. This is a necessary condition to ensure convergence of PDHG (why?).

(Reading this part is optional) To help you understand the context, we briefly summarize the whole procedure:
We are given 2 images $i_0(x, y), i_1(x, y)$ of a scene between which some motion took place. We wish to estimate the motion by evaluating a displacement field $u(x, y) = (u_x(x, y), u_y(x, y))$ such that the warped image $\tilde{i_1}(x, y) = i_1(x+u_x(x, y), y+u_y(x, y))$ matches the reference image i_0. We do so by minimizing the following objective:

$$\int \lambda |i_1(x + u_x, y + u_y) - i_0(x, y)| + \|\nabla u(x, y)\|_2 \, dxdy \tag{37}$$

Linearization of i_1 at $(x + \bar{u}_x, y + \bar{u}_y)$ gives

$$\int \lambda |i_1(x+\bar{u}_x, y+\bar{u}_y) + \langle (u_x-\bar{u}_x, u_y-\bar{u}_y)^\top, \nabla i_1(x+\bar{u}_x, y+\bar{u}_y) \rangle - i_0(x, y)| + \|\nabla u(x, y)\|_2 \, dxdy \tag{38}$$

which is convex in u and can be rewritten as Eq. 38.

In practice, we evaluate U in a coarse-to-fine multi-scale manner where the result at coarser scale serve as an initialization for the optimization at finer scale. Also, for each scale, multiple linearized update steps are performed. These outer loops are given in the template and you only need to focus on solving the inner problem.

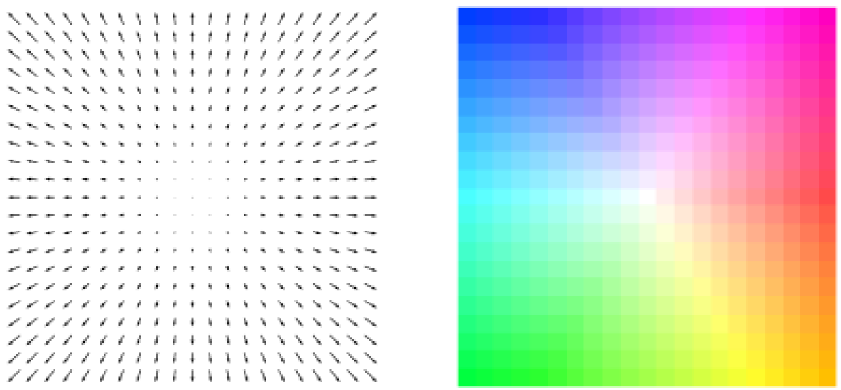

Figure 1: Illustration of optical flow color encoding.

Figure 2: Color encoding of the ground-truth optical flow in the ideal case.