Technische Universität München
Zentrum Mathematik
Lehrstuhl für Mathematische Optimierung, M1
Prof. Dr. Michael Ulbrich
Dennis Klöckner, M.Sc.
Wintersemester 2013/14

First Order Primal-Dual Optimization Methods (Primal-duale Optimierungsmethoden erster Ordnung)

Übungsblatt 3

In der Vorlesung wird folgendes bereitgestellt:

Es sei $f : \mathbb{R}^n \to \mathbb{R}$ eine konvexe Funktion und $x \in \mathbb{R}^n$. Ein Vektor $u \in \mathbb{R}^n$ heißt *Subgradient* von f in x, wenn gilt

$$f(y) \geq f(x) + u^T(y - x) \quad \text{für alle } y \in \mathbb{R}^n.$$

Die Menge der Subgradienten von f in x heißt das (konvexe) *Subdifferential* von f in x und wird mit $\partial f(x)$ bezeichnet.

Es gilt nun, dass $0 \in \partial f(\overline{x})$ genau dann, wenn $\overline{x} = \arg\min_x f(x)$ und wenn f zusätzlich differenzierbar ist, folgt $\partial f(x) = \{\nabla f(x)\}$.

Sei zusätzlich $g : \mathbb{R}^n \to \mathbb{R}$ eine konvexe Funktion. Dann gilt $\partial(f + g) = \partial f + \partial g$.

Aufgabe 3.1 (Shrinkage-Operator):
Wir betrachten die zu minimierende Funktion $\varphi : \mathbb{R} \to \mathbb{R}$ aus der Vorlesung, definiert als

$$\varphi(s) = |s| + \frac{1}{2\nu}(s - t)^2,$$

wobei $t \in \mathbb{R}$ ein Parameter und $\nu > 0$ fest ist. Zeigen Sie zunächst, dass für $h : \mathbb{R} \to \mathbb{R}$ mit $h(s) = |s|$ folgt, dass

$$\partial h(s) = \begin{cases} -1, & \text{falls } s < 0, \\ [-1, 1], & \text{falls } s = 0, \\ 1, & \text{falls } s > 0. \end{cases}$$

Nutzen Sie dies um zu zeigen, dass

$$\arg\min_{s \in \mathbb{R}} \varphi(s) = \operatorname{prox}_{\nu h}(t) = S_\nu(t) := \begin{cases} t + \nu, & \text{falls } t < -\nu, \\ 0, & \text{falls } |t| \leq \nu, \\ t - \nu, & \text{falls } t > \nu, \end{cases}$$

gilt, indem Sie die verschiedenen Fälle von s betrachten und die Eigenschaften des Subdifferentials ausnutzen. Hierbei bezeichnet prox den Proximity-Operator aus Aufgabe 2.1. Der stückweise lineare Operator S_ν heißt *Shrinkage-Operator*.

Aufgabe 3.2 (Proximity-Operator von $\|\cdot\|_2$):
Sei $h : \mathbb{R}^n \to \mathbb{R}$ definiert als $h(x) := \|x\|_2$. Das Subdifferential ist dann (siehe Vorlesung) gegeben durch

$$\partial h(x) = \begin{cases} \{x/\|x\|_2\}, & \text{falls } x \neq 0, \\ \{z \in \mathbb{R}^n : \|z\|_2 \leq 1\}, & \text{falls } x = 0. \end{cases}$$

Zeigen Sie damit, dass der Proximity Operator von τh für $\tau > 0$ gegeben ist durch

$$\operatorname{prox}_{\tau h}(x) = \max\{\|x\|_2 - \tau, 0\} \cdot \frac{x}{\|x\|_2} = \begin{cases} \frac{\|x\|_2 - \tau}{\|x\|_2}\, x, & \text{falls } \|x\|_2 > \tau, \\ 0, & \text{falls } \|x\|_2 \leq \tau, \end{cases}$$

wobei wir die Konvention $0 \cdot 0/0 = 0$ verwenden.

Aufgabe 3.3 (Bildverarbeitungsmodell mit ADMM):
Wir betrachten das (leicht modifizierte) Bildverarbeitungsmodell aus Aufgabe 1.2

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^{n} \|D_i x\|_2 + \mu \|Bx - u\|_1 \quad \text{s.t.} \quad x \in [0,1]^n, \tag{1}$$

mit dem rekonstruierten Bild $x \in \mathbb{R}^n$, bestehend aus n Pixeln, die passend in einem Vektor angeordnet sind. $B : \mathbb{R}^n \to \mathbb{R}^m$ ist ein linearer Operator, $u \in \mathbb{R}^m$ beschreibt die gestörten Daten, $D_i : \mathbb{R}^n \to \mathbb{R}^2$ stehen für diskrete Differenzenoperatoren für alle $i = 1, \ldots, n$ und $\mu > 0$ ist ein Gewichtungsfaktor.

a) Reformulieren Sie (1) als äquivalentes Problem, indem Sie die zusätzlichen Variablen $y_i = D_i x \in \mathbb{R}^2$ für alle $i = 1, \ldots, n$, $z = Bx - u \in \mathbb{R}^m$, und $w = x \in \mathbb{R}^n$ einführen und die Box constraints für w fordern. Weiter sei der Vektor $y \in \mathbb{R}^{2n}$ definiert als $y = (y_1^T, \ldots, y_n^T)^T$.

   Stellen Sie die augmentierte Lagrange-Funktion $L^a(\gamma; x, y, z, w, \lambda, \xi, \zeta)$ auf, wobei $\lambda \in \mathbb{R}^{2n}$, $\xi \in \mathbb{R}^m$ und $\zeta \in \mathbb{R}^n$ die passenden Lagrangemultiplikatoren zu den Gleichungsnebenbedingungen sind, und verwenden Sie $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ als Penalty Parameter mit $\gamma_1, \gamma_2, \gamma_3 > 0$.

b) Wir bezeichnen nun exemplarisch $\min_y L^a(\gamma; x, y, z, w, \lambda, \xi, \zeta)$ als das y-Teilproblem. Formulieren und lösen Sie analog das

   (i) y-Teilproblem, (ii) z-Teilproblem, (iii) w-Teilproblem und (iv) x-Teilproblem.

   Begründen Sie jeweils, dass die Lösungen eindeutig sind.