Relations

Monotone Operators

Fixed Point Iterations

Proximal Point Algorithm

Chapter 5

Operator Splitting Methods

Convex Optimization for Computer Vision

SS 2016

Michael Moeller
Thomas Möllenhoff
Emanuel Laude
Computer Vision Group
Department of Computer Science
TU München

Recap and Motivation

- **Last 3 lectures**: PDHG method for minimizing structured convex problems

$$\min_{u \in \mathbb{R}^n} \; G(u) + F(Ku)$$

- Unintuitive overrelaxation, rather involved convergence analysis
- Next lectures: simple and unified convergence analysis of many different algorithms within a single approach
- Key ideas: monotone operators, fixed point iterations
- Give a new understanding of convex optimization algorithms

Relations

Monotone Operators

Fixed Point Iterations

Proximal Point Algorithm

Relations

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Relations

Monotone Operators

Fixed Point Iterations

Proximal Point Algorithm

Notation

- A relation R on $\mathbb{R}^n$ is a subset of $\mathbb{R}^n \times \mathbb{R}^n$
- We will refer to it as a set-valued **operator** and overload the usual matrix notation

$$R(x) = Rx := \{y \in \mathbb{R}^n \mid (x, y) \in R\}.$$

- If Rx is a singleton or empty for all x, then R is a function (or single-valued operator) with domain

$$\text{dom}(R) := \{x \in \mathbb{R}^n \mid Rx \neq \emptyset\}$$

- Abuse of notation: identify singleton $\{x\}$ with x, i.e., write $Rx = y$ instead of $Rx \ni y$ if R is function
- Concept: identifying functions with their *graph*

Relations

Monotone Operators

Fixed Point Iterations

Proximal Point Algorithm

Some Examples

- Empty relation: $\emptyset$
- Identity: $I := \{(u, u) \mid u \in \mathbb{R}^n\}$
- Zero: $0 := \{(u, 0) \mid u \in \mathbb{R}^n\}$
- Gradient relation:

$$\nabla E := \{(u, \nabla E(u)) \mid u \in \mathbb{R}^n\}$$

- Subdifferential relation:

$$\partial E := \{(u, g) \mid u \in \text{dom}(E), E(v) \geq E(u) + \langle g, v - u \rangle, \forall v \in \mathbb{R}^n\}$$

- Another possible view: think of relations as a set valued functions, e.g., $\partial E : \mathbb{R}^n \rightarrow \mathcal{P}(\mathbb{R}^n)$

Relations

Monotone Operators

Fixed Point Iterations

Proximal Point Algorithm

Our Goal

Solve generalized equation (inclusion) problem

$$0 \in R(u)$$

i.e., find $u \in \mathbb{R}^n$ such that $(u, 0) \in R$.

Examples:

- Set $R = \partial E$, then the goal is to find $0 \in \partial E(u)$
- This are just the optimality conditions of our prototypical optimization problem:

$$\arg\min_{u \in \mathbb{R}^n} \; E(u)$$

- Finding saddle-points $(\tilde{u}, \tilde{p})$ of

$$PD(u,p) = G(u) - F^*(p) + \langle Ku, p \rangle$$

corresponds to the inclusion problem

$$0 \in \begin{bmatrix} \partial G & K^T \\ -K & \partial F^* \end{bmatrix} \begin{bmatrix} u \\ p \end{bmatrix}$$

Operations on Relations

- Inverse $R^{-1} = \{(y, x) \mid (x, y) \in R\}$
  - Exists for *any* relation
  - Reduces to inverse function when R is injective function
- Addition $R + S = \{(x, y + z) \mid (x, y) \in R, (x, z) \in S\}$
- Scaling $\lambda R = \{(x, \lambda y) \mid (x, y) \in R\}$
- Resolvent $J_{\lambda R} := (I + \lambda R)^{-1}$

Examples:

- $I + \lambda R = \{(x, x + \lambda y) \mid (x, y) \in R\}$
- $J_R = \{(x + \lambda y, x) \mid (x, y) \in R\}$
- E closed, proper, convex: $(\partial E)^{-1} = \partial E^*$

$\rightarrow$ **Draw a picture for** $E(u) = |u|$

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Relations

Monotone Operators

Fixed Point Iterations

Proximal Point Algorithm

Monotone Operators

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Relations

Monotone Operators

Fixed Point Iterations

Proximal Point Algorithm

Monotone Operators

Definition

The set-valued operator $T \subset \mathbb{R}^n \times \mathbb{R}^n$ is called **monotone** if

$$\langle u - v, Tu - Tv \rangle \geq 0, \; \forall u, v \in \mathbb{R}^n.$$

An operator T is called **maximally monotone** if it is not contained in any other monotone operator.

- Maximal monotonicity is an important technical detail, but we will be sloppy about it for the rest of the course

Examples of monotone operators:

- Monotonically non-decreasing functions $T : \mathbb{R} \to \mathbb{R}$
- Any positive semi-definite matrix A: $\langle Ax - Ay, x - y \rangle \geq 0$
- Subdifferential of a convex function ∂f
- Proximity operators of convex functions $\text{prox}_{\tau,f} : \mathbb{R}^n \to \mathbb{R}^n$

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Relations

Monotone Operators

Fixed Point Iterations

Proximal Point Algorithm

Monotone Operators

Calculus rules (exercise):

- T monotone, $\lambda \geq 0 \Rightarrow \lambda T$ monotone
- T monotone $\Rightarrow T^{-1}$ monotone
- R, S monotone, $\lambda \geq 0 \Rightarrow R + \lambda S$ is monotone

Some important definitions/properties:

- Lipschitz operators (and in particular nonexpansive operators) are single-valued (functions)
- x is called *fixed point* of operator T if $x = Tx$
- If T is nonexpansive (Lipschitz constant $L \leq 1$) and $\operatorname{dom} T = \mathbb{R}^n$ then the set of fixed points $(I - F)^{-1}(0)$ is closed and convex **(exercise)**

Resolvent and Cayley Operators

- Let $T \subset \mathbb{R}^n \times \mathbb{R}^n$ be set-valued operator
- The *resolvent operator* of T is given as $J_{\lambda T} := (I + \lambda T)^{-1}$
- Special case: $T = \partial f$, $J_{\lambda \partial f}$ is proximal operator of f
- From previous slide: resolvent is monotone if T is monotone
- The *Cayley operator* (or reflection operator) of T is defined as $C_{\lambda T} := 2J_{\lambda T} - I$

Facts:

- $0 \in Tx$ if and only if $x = J_{\lambda T}x = C_{\lambda T}x$
- If T is monotone, then $J_{\lambda T}$ and $C_{\lambda T}$ are nonexpansive

Relations

Monotone Operators

Fixed Point Iterations

Proximal Point Algorithm

Fixed Point Iterations

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Relations

Monotone Operators

Fixed Point Iterations

Proximal Point Algorithm

The Main Algorithm

- Recall that $u \in \mathbb{R}^n$ is fixed point of $T : \mathbb{R}^n \to \mathbb{R}^n$, if $u = Tu$
- The main algorithm of this chapter is the *fixed point* or *Picard iteration* for some given $u^0 \in \mathbb{R}^n$:

$$u^{k+1} = Tu^k, \qquad k = 0, 1, 2, \dots$$

- We will see that many important convex optimization algorithms can be written in this form
- Allows simple and unified analysis

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Relations

Monotone Operators

Fixed Point Iterations

Proximal Point Algorithm

Iteration of Contraction Mappings

Contraction Mapping Theorem

Suppose that $T : \mathbb{R}^n \to \mathbb{R}^n$ is a contraction with Lipschitz constant $L < 1$. Then the fixed point iteration

$$u^{k+1} = Tu^k,$$

also called contraction mapping algorithm, converges to the unique fixed point of T.

$\to$ Proof: see literature[1]

- Example: the gradient method can be written as

$$u^{k+1} = (I - \tau \nabla E)u^k$$

- Suppose E is m-strongly convex and L-smooth, then $I - \tau \nabla E$ is Lipschitz with $L_{GM} = \max\{|1 - \tau m|, |1 - \tau L|\}$
- $I - \tau \nabla E$ is contractive for $\tau \in (0, 2/L)$

[1] This theorem is also known as the Banach fixed point theorem.

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Iteration of Averaged Nonexpansive Mappings

- Recall that a mapping $T : \mathbb{R}^n \to \mathbb{R}^n$ is called *nonexpansive* if it is Lipschitz with constant $L \leq 1$.
- Fixed point iteration of nonexpansive mapping doesn't necessarily converge (example: rotation, reflection)
- The mapping $T : \mathbb{R}^n \to \mathbb{R}^n$ is called *averaged* if $T = (1-\theta)I + \theta N$, for some nonexpansive mapping N and $\theta \in (0, 1)$

Theorem: Krasnosel'skii-Mann

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be averaged, and denote the (non-empty) set of fixed points of T as U. Then the sequence (u^k) produced by the iteration

$$u^{k+1} = Tu^k$$

converges to a fixed point $u^* \in U$, i.e., $u^k \to u^*$.

$\to$ Proof: board!