Chapter 3

Duality

Convex Optimization for Computer Vision

SS 2016

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

Duality

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

updated 10.05.2016

Summary: descent methods

For energies of the form

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

for $E : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ proper, closed, convex, we discussed

Gradient descent:

- $\operatorname{dom} E = \mathbb{R}^n$
- For $E \in \mathcal{F}_L^{1,1}(\mathbb{R}^n)$ energy convergence in $\mathcal{O}(1/k)$
- For $E \in \mathcal{S}_{m,L}^{1,1}(\mathbb{R}^n)$ energy and iterate convergence in $\mathcal{O}(c^k)$

Subgradient descent:

- $\operatorname{dom}(E) = \mathbb{R}^n$
- Applicable to any Lipschitz-continuous convex energy
- Usually rather slow

Gradient projection: Generalizes gradient descent to arbitrary (nonempty, closed, convex) $\operatorname{dom}(E)$.

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

How powerful is the gradient projection algorithm?

Consider the total variation denoising problem

$$u^* \in \operatorname*{argmin}_u \frac{1}{2}\|u - f\|_2^2 + \alpha\|Du\|_{2,1},$$

with the finite difference operator $D : \mathbb{R}^{n\times m\times c} \to \mathbb{R}^{nm\times 2c}$.

Is subgradient descent really the best we can do despite the "nice" strongly convex energy?

Let's try something crazy to try to find a better algorithm:

$$\|g\| = \max_{|q|\leq 1} \langle q, g\rangle$$

Following the crazy idea...

The previous simple observation tells us that

$$\begin{aligned}\|g\|_{2,1} &= \sum_i \|g_i\| = \sum_i \max_{|q_i|\le 1} \langle q_i, g_i\rangle \\ &= \max_{|q_i|\le 1} \underbrace{\sum_i \langle q_i, g_i\rangle}_{=:\langle g,q\rangle} \\ &= \max_{\max_i \|q_i\|\le 1} \langle g, q\rangle = \max_{\|q\|_{2,\infty}\le 1} \langle g, q\rangle\end{aligned}$$

We may write

$$\begin{aligned}\min_u \frac{1}{2}\|u-f\|_2^2 + \alpha\|Du\|_1 &= \min_u \frac{1}{2}\|u-f\|_2^2 + \alpha \max_{\|q\|_{2,\infty}\le 1} \langle Du, q\rangle \\ &= \min_u \max_{\|q\|_{2,\infty}\le 1} \frac{1}{2}\|u-f\|_2^2 + \alpha\langle Du, q\rangle\end{aligned}$$

Can we switch min and max?

TV Minimization

Saddle point problems[1]

Let C and D be non-empty closed convex sets in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, and let S be a continuous finite concave-convex function on $C \times D$. If either C or D is bounded, one has

$$\inf_{v \in D} \sup_{q \in C} S(v, q) = \sup_{q \in C} \inf_{v \in D} S(v, q).$$

We can therefore compute

$$\begin{aligned}\min_u \frac{1}{2}\|u - f\|_2^2 + \alpha\|Du\|_1 &= \min_u \max_{\|q\|_{2,\infty} \leq 1} \frac{1}{2}\|u - f\|_2^2 + \alpha\langle Du, q\rangle \\ &= \max_{\|q\|_{2,\infty} \leq 1} \min_u \frac{1}{2}\|u - f\|_2^2 + \alpha\langle Du, q\rangle\end{aligned}$$

[1] Rockafellar, Convex Analysis, Corollary 37.3.2

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

TV Minimization

Now the inner minimization problem obtains its optimum at

$$
\begin{aligned}
0 &= u - f + \alpha D^* q, \\
\Rightarrow u &= f - \alpha D^* q.
\end{aligned}
$$

The remaining problem in q becomes

$$
\begin{aligned}
&\max_{\|q\|_{2,\infty} \leq 1} \frac{1}{2} \|f - \alpha D^* q - f\|_2^2 + \alpha \langle D(f - \alpha D^* q), q \rangle \\
= &\max_{\|q\|_{2,\infty} \leq 1} \frac{1}{2} \|\alpha D^* q\|_2^2 + \alpha \langle Df, q \rangle - \|\alpha D^* q\|_2^2 \\
= &\max_{\|q\|_{2,\infty} \leq 1} -\frac{1}{2} \|\alpha D^* q - f\|_2^2
\end{aligned}
$$

TV Minimization

Since we prefer minimizations over maximizations, we write

$$\hat{q} = \operatorname*{argmax}_{\|q\|_{2,\infty} \leq 1} -\frac{1}{2}\|\alpha D^* q - f\|_2^2$$
$$= \operatorname*{argmin}_{\|q\|_{2,\infty} \leq 1} \frac{1}{2}\left\|D^* q - \frac{f}{\alpha}\right\|_2^2$$

This is a problem we know how to solve! An L-smooth function over a simple convex set: Gradient projection

$$q^{k+1} = \pi_C\left(q^k - \tau D\left(D^* q^k - \frac{f}{\alpha}\right)\right),$$

where $C = \{q \in \mathbb{R}^{nm \times 2c} \mid \|q\|_{2,\infty} \leq 1\}$.

A conceptual way to reformulate energy minimization problems?

Maybe our idea

$$\|g\| = \max_{|q|\leq 1} \langle q, g \rangle$$

was not so crazy but rather conceptual?

Definition: Convex Conjugate

We define the *convex conjugate* of the function $E : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ to be

$$E^*(p) = \sup_{u \in \mathbb{R}^n} \left(\langle u, p \rangle - E(u) \right).$$

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Convex conjugates

Convexity of the Convex Conjugate

The convex conjugate E^* of any proper function $E : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is convex.

Proof: Board

Convexity of the Convex Conjugate

The convex conjugate E^* of any proper function $E : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is closed.

Proof: Linear functions are closed and arbitrary intersections of closed sets are closed.

Convex conjugates rules

- **Scalar multiplication** :

$$E(u) = \alpha \tilde{E}(u) \;\Rightarrow\; E^*(p) = \alpha \tilde{E}^*(p/\alpha)$$

- **Separable sum**:

$$E(u_1, u_2) = E_1(u_1) + E_2(u_2) \;\Rightarrow\; E^*(p_1, p_2) = E_1^*(p_1) + E_2^*(p_2)$$

- Careful: Only separable sums work this way!
  Sum rule for E_1, E_2 closed, convex, proper:

$$E(u) = E_1(u) + E_2(u) \;\Rightarrow\; E^*(p) = \inf_{p=p_1+p_2} E_1^*(p_1) + E_2^*(p_2).$$

- **Translation**:

$$E(u) = \tilde{E}(u - b) \;\Rightarrow\; E^*(p) = \tilde{E}^*(p) + \langle p, b \rangle$$

- **Additional affine functions**:

$$E(u) = \tilde{E}(u) + \langle b, u \rangle + a \;\Rightarrow\; E^*(p) = \tilde{E}^*(p - b) - a$$

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Convex conjugates

Examples:

- $E(u) = \frac{1}{2}u^2$ leads to $E^*(p) = \frac{1}{2}p^2$
- $E(u) = \|u\|_2$ leads to $E^*(p) = \begin{cases} 0 & \text{if } \|p\|_2 \leq 1, \\ \infty & \text{else.} \end{cases}$
- $E(u) = \|u\|_1$ leads to $E^*(p) = \begin{cases} 0 & \text{if } \|p\|_\infty \leq 1, \\ \infty & \text{else.} \end{cases}$
- $E(u) = \|u\|_\infty$ leads to $E^*(p) = \begin{cases} 0 & \text{if } \|p\|_1 \leq 1, \\ \infty & \text{else.} \end{cases}$
- $E(u) = \begin{cases} 0 & \text{if } \|u\|_2 \leq 1, \\ \infty & \text{else.} \end{cases}$ leads to $E^*(p) = \|p\|_2$.
- $E(u) = \begin{cases} 0 & \text{if } \|u\|_\infty \leq 1, \\ \infty & \text{else.} \end{cases}$ leads to $E^*(p) = \|p\|_1$.
- $E(u) = \begin{cases} 0 & \text{if } \|u\|_1 \leq 1, \\ \infty & \text{else.} \end{cases}$ leads to $E^*(p) = \|p\|_\infty$.

Suspicion: $E^{**} = E$?

Fenchel-Young Inequality

Fenchel-Young Inequality[2]

Let E be proper, convex and closed, $u \in \mathrm{dom}(E) \subset \mathbb{R}^n$, and $p \in \mathbb{R}^n$, then

$$E(u) + E^*(p) \geq \langle u, p \rangle.$$

Equality holds if and only if $p \in \partial E(u)$.

Proof: Board.

[2] Borwein, Zhu *Techniques of variational analysis*, Proposition 4.4.1

Biconjugate

Theorem: Biconjugate[3]

Let $E : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be proper, convex and closed, then $E^{**} = E$.

First incomplete proof on the board.

For the full proof, quickly recall the geometric interpretation of the subgradient of chapter 1, and geometrically convince yourself that the separating hyperplane theorem makes sense.

[3] Rockafellar, Convex Analysis, Theorem 12.2

Geometric interpretation of subgradients

Geometric interpretation of subgradients:

Any subgradient $p \in \partial E(u)$ represents a non-vertical supporting hyperplane to epi(E) at $(u, E(u))$.

Definition

A supporting hyperplane to a set $S \subset \mathbb{R}^n$ is a hyperplane $\{x \in \mathbb{R}^n \mid \langle a, x\rangle = b\}$, such that

- $S \subset \{x \in \mathbb{R}^n \mid \langle a, x\rangle \leq b\}$ or $S \subset \{x \in \mathbb{R}^n \mid \langle a, x\rangle \geq b\}$
- $\exists y \in \partial S$ (the boundary of S) such that $\langle a, y\rangle = b$.

Let $p \in \partial E(u)$. Then

$$
\begin{aligned}
& E(v) - E(u) - \langle p, v - u\rangle \geq 0 && \forall v \in \mathbb{R}^n \\
\Rightarrow\quad & \alpha - E(u) - \langle p, v - u\rangle \geq 0 && \forall (v, \alpha) \in \text{epi}(E) \\
\Rightarrow\quad & \left\langle \begin{bmatrix} -p \\ 1 \end{bmatrix}, \begin{bmatrix} v \\ \alpha \end{bmatrix} - \begin{bmatrix} u \\ E(u) \end{bmatrix} \right\rangle \geq 0 && \forall (v, \alpha) \in \text{epi}(E).
\end{aligned}
$$

Back to the biconjugate

Separating hyperplane theorem

Let $S \subset \mathbb{R}^n$ be a nonempty closed convex set, and $\mathbb{R}^n \ni u \notin S$. Then there exists a nonzero vector z and a number $c < 0$ such that

$$\langle z, v - u \rangle \leq c \qquad \forall v \in S.$$

Theorem: Biconjugate[4]

Let $E : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be proper, convex and closed, then $E^{**} = E$.

Proof: Board.

Now we understand what we did for TV minimization: Replace $\|Du\|_{2,1}$ by

$$(\|\cdot\|_{2,1})^{**}(Du) = \sup_p \ \langle p, Du \rangle - \iota_{\|\cdot\|_{2,\infty} \leq 1}(p).$$

[4] Rockafellar, Convex Analysis, Theorem 12.2

Convex conjugation

Theorem: Subgradient of convex conjugate[5]

Let E be proper, convex and closed, then the following two conditions are equivalent:

- $p \in \partial E(u)$
- $u \in \partial E^*(p)$

Proof: Board

Board: A quick way for repeating our TV-reformulation.

Conjugation of strongly convex functions

If $E : \mathbb{R}^n \to \bar{\mathbb{R}}$ is proper, closed and m-strongly convex, then $E^* \in \mathcal{F}^{1,1}_{\frac{1}{m}}(\mathbb{R}^n)$.

Proof: Board

[5] Rockafellar, Convex Analysis, Theorem 23.5

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Conjugation with linear operators

Definition: Image function

Let $A :\in \mathbb{R}^{n\times m}$ be a linear operator and $E : \mathbb{R}^m \to \bar{\mathbb{R}}$ an extended real valued function. We call $AE : \mathbb{R}^n \to \bar{\mathbb{R}}$ defined via

$$AE(u) := \begin{cases} \inf_{v, Av=u} E(v) & \text{if } u \in \text{range}(A), \\ \infty & \text{else.} \end{cases}$$

the *image of E under A*.

Conjugate of image functions

Let E be a proper extended real valued function and let A be a linear operator. Then it holds that

$$(AE)^* = (E^* \circ A^*).$$

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Conjugation with linear operators

Conjugation with linear operators

Let E be a proper closed convex function and let A be a linear operator such that

$$\mathrm{im}(A) \cap \mathrm{ri}(\mathrm{dom}(E)) \neq \emptyset.$$

Then

$$(E^* \circ A^*)^* = AE,$$

Proof: Not carried out, but key is that the above assumption ensures that AE is proper and closed such that $(AE)^{**} = AE$.

For details on the question of why $\mathrm{im}(A) \cap \mathrm{ri}(\mathrm{dom}(E)) \neq \emptyset$ is sufficient for the closedness of AE you can consider "Fundamentals of Convex Analysis" by Jean-Baptiste Hiriart-Urruty, Claude Lemarechal, pp. 224–225, Lemma 2.2.2

Fenchel duality

Fenchel's Duality Theorem[6]

Let $H : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ and $R : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ be proper, closed, convex functions and let there exists a $u \in \mathrm{ri}(\mathrm{dom}(H))$ such that $Ku \in \mathrm{ri}(\mathrm{dom}(R))$. Then

$$\boxed{\begin{array}{lll} & \inf_u & H(u) + R(Ku) & \text{"Primal"} \\ = & \inf_u \sup_q & H(u) + \langle q, Ku \rangle - R^*(q) & \text{"Saddle point"} \\ = & \sup_q \inf_u & H(u) + \langle q, Ku \rangle - R^*(q) & \\ = & \sup_q & -H^*(-K^* q) - R^*(q) & \text{"Dual"} \end{array}}$$

Proof: Board.

[6] C.f. Rockafellar, *Convex Analysis*, Section 31

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Applications and motivation of the next lecture

Assume we want to minimize

$$\min_u \frac{1}{2}\|u - f\|_2^2 \text{ s.t. } \|Du\|_\infty \leq c,$$

i.e. find the best approximation to the input data f under the constraint that Du must be bounded componentwise.

Dual problem:

$$\max_p -\frac{1}{2}\|D^* p\|^2 - \langle p, f\rangle - c\|p\|_1$$

or

$$\hat{p} = \operatorname*{argmin}_p \frac{1}{2}\|D^* p + f\|^2 + c\|p\|_1$$

We have seen how to solve such a problem by using $p = p_1 - p_2$ with $p_1, p_2 \geq 0$. **Next two lectures: Solution without doubling the number of free variables!**

Michael Moeller
Thomas Möllenhoff
Emanuel Laude