Chapter 3
Duality

Convex Optimization for Computer Vision
SS 2019

Michael Moeller
Chair for Computer Vision
University of Siegen

Duality

Summary: descent methods

For energies of the form

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

for proper, closed, convex $F : \mathbb{R}^n \to \mathbb{R}$, $G : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, with F additionally being L-smooth, we discussed

Gradient descent: $G \equiv 0$

Gradient projection: $G = \delta_C$

Proximal gradient: G simple (easy to compute prox)

Convergence rates

- Energy convergence in $\mathcal{O}(1/k)$ for "plain" method
- Energy convergence in $\mathcal{O}(1/k^2)$ for Nesterov's method
- Energy and iterate convergence in $\mathcal{O}(c^k)$, $c < 1$, for strongly convex energies.

How powerful is the gradient projection algorithm?

Consider the total variation denoising problem

$$u^* \in \underset{u}{\operatorname{argmin}} \frac{1}{2}\|u - f\|_2^2 + \alpha \|Du\|_1,$$

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

How powerful is the gradient projection algorithm?

Consider the total variation denoising problem

$$u^* \in \operatorname*{argmin}_u \frac{1}{2}\|u - f\|_2^2 + \alpha\|Du\|_1,$$

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?

How powerful is the gradient projection algorithm?

Consider the total variation denoising problem

$$u^* \in \underset{u}{\operatorname{argmin}} \frac{1}{2}\|u - f\|_2^2 + \alpha \|Du\|_1,$$

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:

$$|g| = \max_{|q| \leq 1} q \cdot g$$

Following the crazy idea...

The previous simple observation tells us that

$$\|g\|_1 = \sum_i |g_i| = \sum_i \max_{|q_i|\leq 1} q_i \cdot g_i$$

$$= \max_{|q_i|\leq 1, \forall i} \underbrace{\sum_i q_i \cdot g_i}_{=:\langle g, q\rangle}$$

$$= \max_{\max_i |q_i|\leq 1} \langle g, q\rangle = \max_{\|q\|_\infty \leq 1} \langle g, q\rangle$$

Following the crazy idea...

The previous simple observation tells us that

$$\|g\|_1 = \sum_i |g_i| = \sum_i \max_{|q_i|\leq 1} q_i \cdot g_i$$

$$= \max_{|q_i|\leq 1, \forall i} \underbrace{\sum_i q_i \cdot g_i}_{=:\langle g, q\rangle}$$

$$= \max_{\max_i |q_i|\leq 1} \langle g, q\rangle = \max_{\|q\|_\infty \leq 1} \langle g, q\rangle$$

We may write

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

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

Following the crazy idea...

The previous simple observation tells us that

$$\|g\|_1 = \sum_i |g_i| = \sum_i \max_{|q_i|\leq 1} q_i \cdot g_i$$

$$= \max_{|q_i|\leq 1, \forall i} \underbrace{\sum_i q_i \cdot g_i}_{=:\langle g,q\rangle}$$

$$= \max_{\max_i |q_i|\leq 1} \langle g, q\rangle = \max_{\|q\|_\infty \leq 1} \langle g, q\rangle$$

We may write

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

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

Can we switch min and max?

TV Minimization

Saddle point problems[a]

[a] Rockafellar, Convex Analysis, Corollary 37.3.2

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

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

TV Minimization

Saddle point problems[a]

[a] Rockafellar, Convex Analysis, Corollary 37.3.2

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

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

TV Minimization

Now the inner minimization problem obtains its optimum at

$$0 = u - f + \alpha D^* q,$$
$$\Rightarrow u = f - \alpha D^* q.$$

The remaining problem in q becomes

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

TV Minimization

Since we prefer minimizations over maximizations, we write

$$\hat{q} = \underset{\|q\|_\infty \leq 1}{\operatorname{argmin}} \frac{1}{2} \|\alpha D^* q\|_2^2 - \alpha \langle Df, q \rangle$$

$$= \underset{\|q\|_\infty \leq 1}{\operatorname{argmin}} \frac{1}{2} \|D^* q\|_2^2 - \frac{1}{\alpha} \langle Df, q \rangle$$

TV Minimization

Since we prefer minimizations over maximizations, we write

$$\hat{q} = \underset{\|q\|_\infty \leq 1}{\operatorname{argmin}} \frac{1}{2}\|\alpha D^* q\|_2^2 - \alpha\langle Df, q\rangle$$
$$= \underset{\|q\|_\infty \leq 1}{\operatorname{argmin}} \frac{1}{2}\|D^* q\|_2^2 - \frac{1}{\alpha}\langle Df, q\rangle$$

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

A conceptual way to reformulate energy minimization problems?

Maybe our idea

$$\|g\|_1 = \max_{|q|_\infty \leq 1} \langle q, g \rangle$$

was not so crazy but rather conceptual?

A conceptual way to reformulate energy minimization problems?

Maybe our idea

$$\|g\|_1 = \max_{|q|_\infty \leq 1} \langle q, g \rangle$$

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

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

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 and closed.

Proof: Exercise

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 and closed.

Proof: Exercise

Are there reasonable computation rules for the convex conjugate that simplify our lives in practice?

Convex conjugates rules

- **Scalar multiplication** :

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

Convex conjugates rules

- **Scalar multiplication** :

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

- **Separable sum**:

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

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:

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

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**:

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

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**:

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

Convex conjugates

Examples:

- $E(u) = \frac{1}{2}u^2$ leads to $E^*(p) = \frac{1}{2}p^2$

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 $E^*(p) = \begin{cases} 0 & \text{if } \|p\|_2 \leq 1, \\ \infty & \text{else.} \end{cases}$

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 $E^*(p) = \begin{cases} 0 & \text{if } \|p\|_\infty \leq 1, \\ \infty & \text{else.} \end{cases}$

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 $E^*(p) = \begin{cases} 0 & \text{if } \|p\|_1 \leq 1, \\ \infty & \text{else.} \end{cases}$

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 $E^*(p) = \|p\|_2$.

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 $E^*(p) = \|p\|_1$.

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[a]

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

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

Biconjugate

Theorem: Biconjugate[a]

[a] Rockafellar, Convex Analysis, Theorem 12.2

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

Incomplete proof on the board.

Biconjugate

Theorem: Biconjugate[a]

[a] Rockafellar, Convex Analysis, Theorem 12.2

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

Incomplete proof on the board.

Now we understand what we did for TV minimization: Replace $\|Du\|_1$ by

$$(\|\cdot\|_1)^{**}(Du) = \sup_p \langle p, Du \rangle - \delta_{\|\cdot\|_\infty \le 1}(p).$$

Convex conjugation

Theorem: Subgradient of convex conjugate[a]

[a] Rockafellar, Convex Analysis, Theorem 23.5

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

Convex conjugation

Theorem: Subgradient of convex conjugate[a]

[a] Rockafellar, Convex Analysis, Theorem 23.5

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.

Fenchel duality

Fenchel's Duality Theorem[a]

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

Let $G : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ and $F : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ be proper, closed, convex functions, $K \in \mathbb{R}^{m \times n}$, and let there exist a $u \in \text{ri}(\text{dom}(G))$ such that $Ku \in \text{ri}(\text{dom}(F))$. Then

$$
\begin{aligned}
& \inf_u && G(u) + F(Ku) && \text{"Primal"} \\
= \; & \inf_u \sup_p && G(u) + \langle p, Ku \rangle - F^*(p) && \\
& && && \text{"Saddle point"} \\
= \; & \sup_p \inf_u && G(u) + \langle p, Ku \rangle - F^*(p) && \\
= \; & \sup_p && -G^*(-K^T p) - F^*(p) && \text{"Dual"}
\end{aligned}
$$

Relations between primal and dual variables

Conclusion

Let the assumptions from Fenchel's Duality Theorem hold. If there exists a pair $(u, p) \in \mathbb{R}^n \times \mathbb{R}^n$ such that one of the following four equivalent conditions are met

1. $-K^T p \in \partial G(u), \quad p \in \partial F(Ku),$
2. $-K^T p \in \partial G(u), \quad Ku \in \partial F^*(p),$
3. $u \in \partial G^*(-K^T p), \quad p \in \partial F(Ku),$
4. $u \in \partial G^*(-K^T p), \quad Ku \in \partial F^*(p),$

Then u solves the primal and p solves the dual optimization problem.

Example application of duality

Assume we want to minimize

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

Example application of duality

Assume we want to minimize

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

Dual problem:

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

or

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

Example application of duality

Assume we want to minimize

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

Dual problem:

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

or

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

We can apply the proximal gradient algorithm!

Example application of duality

Assume we want to minimize

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

Dual problem:

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

or

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

We can apply the proximal gradient algorithm!

Knowing in advance if the dual problem is more 'friendly':

Conjugation of strongly convex functions

If $E : \mathbb{R}^n \to \bar{\mathbb{R}}$ is proper, closed and m-strongly convex, then E^* is proper, closed, convex and $1/m$-smooth.

Does this solve all problems?

Consider TV-ℓ^1 denoising, i.e.,

$$\begin{aligned}&\inf_u \|u-f\|_1 + \alpha\|Du\|_1 \\ =&\sup_q \langle \alpha D^* q, f\rangle - \delta_{\|\cdot\|_\infty \leq 1}(-\alpha D^* q) - \delta_{\|\cdot\|_\infty \leq 1}(q)\end{aligned}$$

Does this solve all problems?

Consider TV-ℓ^1 denoising, i.e.,

$$\begin{aligned}&\inf_u \; \|u - f\|_1 + \alpha\|Du\|_1 \\ =&\sup_q \; \langle \alpha D^* q, f\rangle - \delta_{\|\cdot\|_\infty \leq 1}(-\alpha D^* q) - \delta_{\|\cdot\|_\infty \leq 1}(q)\end{aligned}$$

The problem did not become easier! What can we do?

Does this solve all problems?

Consider TV-ℓ^1 denoising, i.e.,

$$\inf_u \; \|u - f\|_1 + \alpha \|Du\|_1$$
$$= \sup_q \; \langle \alpha D^* q, f \rangle - \delta_{\|\cdot\|_\infty \leq 1}(-\alpha D^* q) - \delta_{\|\cdot\|_\infty \leq 1}(q)$$

The problem did not become easier! What can we do?

Next chapter

Work on the saddle-point problem direct! Try to find (u, q) with

$$-K^T q \in \partial G(u), \quad Ku \in \partial F^*(q).$$