Duality

V. Estellers

WS 2017

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

Strongly convex energies, convergence $\mathcal{O}(c^k)$ for energy/iterates.

Limitations of Direct Gradient Projection

Given $D : \mathbb{R}^{n\times m\times c} \to \mathbb{R}^{nm\times 2c}$ the finite difference operator, consider the total variation denoising problem

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

Is subgradient descent the best we can do despite strong convexity?

Let's try to remove D from the $\|\cdot\|_{2,1}$ by re-formulating the energy:

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

Alternative Formulation

The previous simple observation tells us that

$$\|g\|_{2,1} = \sum_i \|g_i\| = \sum_i \max_{|q_i|\leq 1} \langle q_i, g_i \rangle = \max_{|q_i|\leq 1} \sum_i \langle q_i, g_i \rangle = \max_{\|q_i\|\leq 1} \langle g, q \rangle$$

$$\|g\|_{2,1} = \max_{\|q\|_{2,\infty}\leq 1} \langle g, q \rangle$$

We may write

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

Can we switch min and max?

Alternative Formulation of TV Minimization

Theorem (Rockafellar, Convex Analysis, 37.3.2[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

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

[1] This absolutely non-trivial. Consider $S(v,q) = cos(v+q)$, $D = C = \mathbb{R}$.

Alternative Formulation of 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

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

Alternative Formulation of TV Minimization

Since we prefer minimizations over maximizations, we write

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

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

$$\hat{q} = \underset{q\in C}{\operatorname{argmin}} \frac{1}{2}\left\|D^* q - \frac{f}{\alpha}\right\|_2^2 \text{ where } C = \{q \in \mathbb{R}^{nm\times 2c} \mid \|q\|_{2,\infty} \leq 1\}$$

Minimization of a convex, proper, closed, L-smooth function over a convex set C. We can solve it with gradient projection: .

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

Can we Generalize this Reformulation? Conjugation

The key idea of our reformulation is

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

Definition

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

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

Lemma

Convexity of the Convex Conjugate The convex conjugate

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

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

Proof: Not

Convex Conjugate Rules

Scalar multiplication :

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

Separable sum:

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

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:

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

Examples of Convex Conjugates

Do you see a pattern?

$E(u) = \frac{1}{2}u^2$ 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 $E^*(p) = \|p\|_\infty$.

Fenchel-Young Inequality

Theorem (Fenchel-Young Inequality)

Let E be proper, convex and closed, $u \in dom(E) \subset \mathbb{R}^n$, and $p \in \mathbb{R}^n$, then $E(u) + E^*(p) \geq \langle u, p \rangle$. Equality holds if and only if $p \in \partial E(u)$.

Theorem (Biconjugate)

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

For TV minimization we replaced $\|Du\|_{2,1}$ by

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

Convex conjugation

Theorem

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

$$p \in \partial E(u)$$

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

Fenchel duality

Theorem (Fenchel's Duality)

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 and let there exist a* $u \in$ *ri*$($*dom*$(G))$ *such that* $Ku \in$ *ri*$($*dom*$(F))$*. Then*

$$\boxed{\begin{array}{llll} & \inf_u & G(u) + F(Ku) & \textit{"Primal"} \\ = & \inf_u \sup_q & G(u) + \langle q, Ku \rangle - F^*(q) & \\ & & & \textit{"Saddle point"} \\ = & \sup_q \inf_u & G(u) + \langle q, Ku \rangle - F^*(q) & \\ = & \sup_q & -G^*(-K^* q) - F^*(q) & \textit{"Dual"} \end{array}}$$

Relations between primal and dual variables

Corollary

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

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

Then u *solves the primal and* q *solves the dual optimization problem.*

Example application of duality

Assume we want to find the best approximation to the input data f under the constraint that Du must be bounded componentwise

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

The dual problem can be solve with proximal gradient

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

Can we know in advance if the dual problem is more “friendly”?

Theorem

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.,

$$
\begin{aligned}
&\inf_u \; \|u - f\|_1 + \alpha \|Du\|_{2,1} \\
=& \inf_u \sup_q \; \|u - f\|_1 + \alpha \langle q, Du \rangle - \delta_{\|\cdot\|_{2,\infty} \leq 1}(q) \\
=& \sup_q \inf_u \; \|u - f\|_1 + \alpha \langle q, Du \rangle - \delta_{\|\cdot\|_{2,\infty} \leq 1}(q) \\
=& \sup_q \left(- \sup_u \; \langle -\alpha D^* q, u \rangle - \|u - f\|_1 \right) - \delta_{\|\cdot\|_{2,\infty} \leq 1}(q) \\
=& \sup_q \; \langle \alpha D^* q, f \rangle - \delta_{\|\cdot\|_\infty \leq 1}(-\alpha D^* q) - \delta_{\|\cdot\|_{2,\infty} \leq 1}(q)
\end{aligned}
$$

The problem did not become easier. We will see next week that we can work on the saddle-point problem directly.