Iterative Algorithms in Optimization, Variational Analysis and Fixed Point Theory

Unit 05: Duality.

university of groningen

Conjugate functions

The (Fenchel) conjugate of a closed convex function $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is the function $f^* : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ defined by

$$f^*(x^*) = \sup_{x \in \mathbb{R}^N} \{x^* \cdot x - f(x)\}.$$

Conjugate functions

The (Fenchel) conjugate of a closed convex function $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is the function $f^* : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ defined by

$$f^*(x^*) = \sup_{x \in \mathbb{R}^N} \{x^* \cdot x - f(x)\}.$$

Examples

- In $\mathbb{R}$: (1) $f(x) = x$;

Conjugate functions

The (Fenchel) conjugate of a closed convex function $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is the function $f^* : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ defined by

$$f^*(x^*) = \sup_{x \in \mathbb{R}^N} \{x^* \cdot x - f(x)\}.$$

Examples

- In $\mathbb{R}$: (1) $f(x) = x$; (2) $f(x) = \frac{1}{2}x^2$;

Conjugate functions

The (Fenchel) conjugate of a closed convex function $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is the function $f^* : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ defined by

$$f^*(x^*) = \sup_{x \in \mathbb{R}^N} \{x^* \cdot x - f(x)\}.$$

Examples

- In $\mathbb{R}$: (1) $f(x) = x$; (2) $f(x) = \frac{1}{2}x^2$; (3) $f(x) = e^x$.

Conjugate functions

The (Fenchel) conjugate of a closed convex function $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is the function $f^* : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ defined by

$$f^*(x^*) = \sup_{x \in \mathbb{R}^N} \{x^* \cdot x - f(x)\}.$$

Examples

- In $\mathbb{R}$: (1) $f(x) = x$; (2) $f(x) = \frac{1}{2}x^2$; (3) $f(x) = e^x$.
- In $\mathbb{R}^N$: (4) $f(x) = c \cdot x + \alpha$;

Conjugate functions

The (Fenchel) conjugate of a closed convex function $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is the function $f^* : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ defined by

$$f^*(x^*) = \sup_{x \in \mathbb{R}^N} \{x^* \cdot x - f(x)\}.$$

Examples

- In $\mathbb{R}$: (1) $f(x) = x$; (2) $f(x) = \frac{1}{2}x^2$; (3) $f(x) = e^x$.
- In $\mathbb{R}^N$: (4) $f(x) = c \cdot x + \alpha$; (5) $f(x) = \phi(\|x\|)$.

Basic properties

Proposition

1. *The function f^* is closed, convex and not identically $+\infty$.*

Basic properties

Proposition

1. *The function f^* is closed, convex and not identically $+\infty$.*
2. *If $f \leq g$, then $f^* \geq g^*$.*

Basic properties

Proposition

1. *The function f^* is closed, convex and not identically $+\infty$.*
2. *If $f \leq g$, then $f^* \geq g^*$.*
3. *Fenchel-Young Inequality: $f(x) + f^*(x^*) \geq x^* \cdot x$.*

Basic properties

Proposition

1. *The function f^* is closed, convex and not identically $+\infty$.*
2. *If $f \leq g$, then $f^* \geq g^*$.*
3. *Fenchel-Young Inequality:* $f(x) + f^*(x^*) \geq x^* \cdot x$.
   There is equality if, and only if, $x^* \in \partial f(x)$.

Basic properties

Proposition

1. *The function f^* is closed, convex and not identically $+\infty$.*
2. *If $f \leq g$, then $f^* \geq g^*$.*
3. *Fenchel-Young Inequality: $f(x) + f^*(x^*) \geq x^* \cdot x$.*
   There is equality if, and only if, $x^* \in \partial f(x)$.
4. $f^{**} = f$.

Basic properties

Proposition

1. *The function f^* is closed, convex and not identically $+\infty$.*
2. *If $f \leq g$, then $f^* \geq g^*$.*
3. *Fenchel-Young Inequality: $f(x) + f^*(x^*) \geq x^* \cdot x$.*
   There is equality if, and only if, $x^* \in \partial f(x)$.
4. $f^{**} = f$.
5. *Legendre-Fenchel Reciprocity Formula: $x^* \in \partial f(x)$ if, and only if, $x \in \partial f^*(x^*)$.*

Basic properties

Proposition

1. *The function f^* is closed, convex and not identically $+\infty$.*
2. *If $f \leq g$, then $f^* \geq g^*$.*
3. *Fenchel-Young Inequality: $f(x) + f^*(x^*) \geq x^* \cdot x$.*
   There is equality if, and only if, $x^* \in \partial f(x)$.
4. $f^{**} = f$.
5. *Legendre-Fenchel Reciprocity Formula: $x^* \in \partial f(x)$ if, and only if, $x \in \partial f^*(x^*)$.*
6. *Let $\mu\ell = 1$. Then, f is μ-strongly convex if, and only if, f^* is ℓ-smooth.*

Fenchel-Rockafellar duality

Let $P \in \mathbb{R}^{M \times N}$, and let $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ and $g : \mathbb{R}^M \to \mathbb{R} \cup \{+\infty\}$ be closed and convex.

Fenchel-Rockafellar duality

Let $P \in \mathbb{R}^{M \times N}$, and let $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ and $g : \mathbb{R}^M \to \mathbb{R} \cup \{+\infty\}$ be closed and convex.

The primal problem is $\inf_{x \in \mathbb{R}^N} \big\{ f(x) + g(Px) \big\}$, with optimal value $v \in \mathbb{R}$, and set of primal solutions $S \subset \mathbb{R}^N$.

Fenchel-Rockafellar duality

Let $P \in \mathbb{R}^{M \times N}$, and let $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ and $g : \mathbb{R}^M \to \mathbb{R} \cup \{+\infty\}$ be closed and convex.

The primal problem is $\inf_{x \in \mathbb{R}^N} \big\{ f(x) + g(Px) \big\}$, with optimal value $v \in \mathbb{R}$, and set of primal solutions $S \subset \mathbb{R}^N$.

The dual problem is $\inf_{y \in \mathbb{R}^M} \big\{ f^*(-P^T y) + g^*(y) \big\}$, with optimal value $v^* \in \mathbb{R}$, and set of dual solutions $S^* \subset \mathbb{R}^M$.

Fenchel-Rockafellar duality

Let $P \in \mathbb{R}^{M\times N}$, and let $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ and $g : \mathbb{R}^M \to \mathbb{R} \cup \{+\infty\}$ be closed and convex.

The primal problem is $\inf_{x\in\mathbb{R}^N} \big\{ f(x) + g(Px) \big\}$, with optimal value $v \in \mathbb{R}$, and set of primal solutions $S \subset \mathbb{R}^N$.

The dual problem is $\inf_{y\in\mathbb{R}^M} \big\{ f^*(-P^T y) + g^*(y) \big\}$, with optimal value $v^* \in \mathbb{R}$, and set of dual solutions $S^* \subset \mathbb{R}^M$.

Proposition

The duality gap $v + v^*$ is nonnegative.

Characterization of the primal-dual solutions

Theorem

The following statements concerning $\hat{x} \in \mathbb{R}^N$ and $\hat{y} \in \mathbb{R}^M$ are equivalent:

i) $-P^T \hat{y} \in \partial f(\hat{x})$ *and* $\hat{y} \in \partial g(P\hat{x})$;

ii) $f(\hat{x}) + f^*(-P^T \hat{y}) = -P^T \hat{y} \cdot \hat{x}$ *and* $g(P\hat{x}) + g^*(\hat{y}) = \hat{y} \cdot P\hat{x}$;

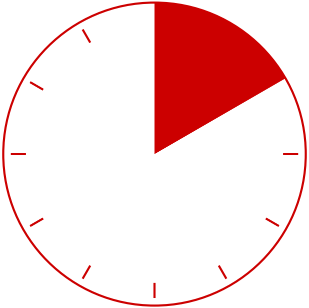
iii) $f(\hat{x}) + g(P\hat{x}) + f^*(-P^T \hat{y}) + g^*(\hat{y}) = 0$; *and*

iv) $\hat{x} \in S$ *and* $\hat{y} \in S^*$ *and* $v + v^* = 0$.

Moreover, if $\hat{x} \in S$ and g is continuous[a], there exists $\hat{y} \in \mathbb{R}^M$ such that all four statements hold.

[a] This can be made much more general.

Break



Structured optimization problem

We consider the problem

$$\min \{f(x) + g(Px) + h(x)\},$$

where

- $P \in \mathbb{R}^{M\times N}$;
- $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ and $g : \mathbb{R}^M \to \mathbb{R} \cup \{+\infty\}$ are closed and convex; and
- $h : \mathbb{R}^N \to \mathbb{R}$ is ℓ-smooth and convex.

Primal-dual algorithm

Chambolle-Pock (2011), Condat-Vũ, (2013):

$$\left\{\begin{array}{rcl} x_{k+1} & = & \mathrm{prox}_{\tau f}\left(x_k - \tau\nabla h(x_k) - \tau P^T y_k\right) \\ y_{k+1} & = & \mathrm{prox}_{\sigma g^*}\left(y_k + \sigma P(2x_{k+1} - x_k)\right), \end{array}\right.$$

with $\tau\sigma\|P\|^2 + \frac{\tau\ell}{2} \leq 1$.

Primal-dual algorithm

Chambolle-Pock (2011), Condat-Vũ, (2013):

$$\begin{cases} x_{k+1} &= \operatorname{prox}_{\tau f}\left(x_k - \tau \nabla h(x_k) - \tau P^T y_k\right) \\ y_{k+1} &= \operatorname{prox}_{\sigma g^*}\left(y_k + \sigma P(2x_{k+1} - x_k)\right), \end{cases}$$

with $\tau\sigma\|P\|^2 + \frac{\tau\ell}{2} \leq 1$.

Remark

Limit points are solutions of the problem.

Primal-dual algorithm

Chambolle-Pock (2011), Condat-Vũ, (2013):

$$\begin{cases} x_{k+1} &= \operatorname{prox}_{\tau f}\left(x_k - \tau\nabla h(x_k) - \tau P^T y_k\right) \\ y_{k+1} &= \operatorname{prox}_{\sigma g^*}\left(y_k + \sigma P(2x_{k+1} - x_k)\right), \end{cases}$$

with $\tau\sigma\|P\|^2 + \frac{\tau\ell}{2} \leq 1$.

Remark

Limit points are solutions of the problem.

Implementation trick: Moreau's Identity

$$\operatorname{prox}_{\sigma g^*}(y) = y - \sigma \operatorname{prox}_{\sigma^{-1}g}\left(\sigma^{-1}y\right).$$

TV Regularization

The Total Variation Regularization Problem is

$$\min_{x\in\mathbb{R}^{N_1\times N_2}} \left\{ \frac{1}{2}\|Fx-b\|^2 + \rho\|Dx\|_1 \right\},$$

where F models or approximates the process by which an image x has been modified (usually deteriorated) to produce b, and D is the discrete gradient.

TV Regularization

The Total Variation Regularization Problem is

$$\min_{x\in\mathbb{R}^{N_1\times N_2}} \left\{ \frac{1}{2}\|Fx-b\|^2 + \rho\|Dx\|_1 \right\},$$

where F models or approximates the process by which an image x has been modified (usually deteriorated) to produce b, and D is the discrete gradient.

Question

Can we apply the primal-dual algorithm to this problem?

Iterative Algorithms in Optimization, Variational Analysis and Fixed Point Theory

Unit 05: Duality.

university of groningen

Reminder I

Terminology and notation for duality

Let $P \in \mathbb{R}^{M\times N}$, and let $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ and $g : \mathbb{R}^M \to \mathbb{R} \cup \{+\infty\}$ be closed and convex.

The primal problem is $\inf_{x\in\mathbb{R}^N} \big\{ f(x) + g(Px) \big\}$, with optimal value $v \in \mathbb{R}$, and set of primal solutions $S \subset \mathbb{R}^N$.

The dual problem is $\inf_{u\in\mathbb{R}^M} \big\{ f^*(-P^T u) + g^*(u) \big\}$, with optimal value $v^* \in \mathbb{R}$, and set of dual solutions $S^* \subset \mathbb{R}^M$.

Linear programming

The linear programming problem is

$$(LP) \qquad \min_{x\in\mathbb{R}^N} \{\, c\cdot x \;:\; Ax \leq b \,\},$$

where $c \in \mathbb{R}^N$, A is a matrix of size $M \times N$, and $b \in \mathbb{R}^M$.

Linear programming

The linear programming problem is

$$(LP) \qquad \min_{x\in\mathbb{R}^N} \{\, c\cdot x \;:\; Ax \leq b \,\},$$

where $c \in \mathbb{R}^N$, A is a matrix of size $M \times N$, and $b \in \mathbb{R}^M$.

It is a primal problem with $f(x) = c\cdot x$ and $g(z) = \iota_{R_+^M}(b - z)$.

Linear programming

The linear programming problem is

$$(LP) \qquad \min_{x\in\mathbb{R}^N} \{\, c\cdot x \;:\; Ax \le b \,\},$$

where $c \in \mathbb{R}^N$, A is a matrix of size $M \times N$, and $b \in \mathbb{R}^M$.

It is a primal problem with $f(x) = c \cdot x$ and $g(z) = \iota_{R_+^M}(b - z)$.

Dual problem

$$(DLP) \qquad \min_{u\in\mathbb{R}^M} \{\, b\cdot u \;:\; A^T u + c = 0, \text{ and } u \ge 0 \,\}.$$

Linear programming

The linear programming problem is

$$(LP) \qquad \min_{x\in\mathbb{R}^N} \{ c\cdot x \;:\; Ax \le b \},$$

where $c \in \mathbb{R}^N$, A is a matrix of size $M \times N$, and $b \in \mathbb{R}^M$.

It is a primal problem with $f(x) = c \cdot x$ and $g(z) = \iota_{R^M_+}(b - z)$.

Dual problem

$$(DLP) \qquad \min_{u\in\mathbb{R}^M} \{ b\cdot u \;:\; A^T u + c = 0, \text{ and } u \ge 0 \}.$$

Exercise

1. Compute the conjugates. Show that (DLP) is the dual of (LP).
2. Compute the dual of the dual.

Reminder II

Highlights of the Duality Theorem

Theorem

The following statements concerning $\hat{x} \in \mathbb{R}^N$ and $\hat{u} \in \mathbb{R}^M$ are equivalent:

i) $-P^T\hat{u} \in \partial f(\hat{x})$ *and* $\hat{u} \in \partial g(P\hat{x})$;

ii) $\hat{x} \in S$ *and* $\hat{u} \in S^*$ *and* $v + v^* = 0$.

Moreover, if $\hat{x} \in S$ and g is continuous[a]*, there exists $\hat{u} \in \mathbb{R}^M$ such that all four statements hold.*

[a] This can be made much more general.

Linearly constrained problems

Consider the problem

$$\min \{f(x) : Px = b\}$$

Linearly constrained problems

Consider the problem

$$\min \{f(x) : Px = b\},$$

whose dual is

$$\min \left\{ f^*(-P^T u) + b \cdot u \right\}.$$

Linearly constrained problems

Consider the problem

$$\min \{f(x) : Px = b\},$$

whose dual is

$$\min \left\{ f^*(-P^T u) + b \cdot u \right\}.$$

Every primal-dual solution $(\hat{x}, \hat{u})$ must satisfy

$$-P^T \hat{u} \in \partial f(\hat{x}) \quad \text{and} \quad P\hat{x} = b.$$

Linearly constrained problems

Consider the problem

$$\min \{f(x) : Px = b\},$$

whose dual is

$$\min \left\{ f^*(-P^T u) + b \cdot u \right\}.$$

Every primal-dual solution $(\hat{x}, \hat{u})$ must satisfy

$$-P^T \hat{u} \in \partial f(\hat{x}) \quad \text{and} \quad P\hat{x} = b.$$

In other words, it is a critical point of the Lagrangian

$$L(x, u) = f(x) + u^T (Px - b).$$

Exercise

Verify all this.

Lagrangian Algorithm

We generate a sequence (x_k, u_k) by iterating

$$\begin{cases} x_{k+1} \in \operatorname{Argmin}\left\{ f(x) + u_k^T(Px - b) \ : \ x \in \mathbb{R}^N \right\}^1, \\ u_{k+1} = u_k + \gamma\big(Px_{k+1} - b\big), \quad \text{with} \quad \gamma > 0. \end{cases}$$

[1] Assuming this has a solution. When is it the case?

Lagrangian Algorithm

We generate a sequence (x_k, u_k) by iterating

$$\begin{cases} x_{k+1} \in \operatorname{Argmin}\left\{ f(x) + u_k^T(Px - b) \;:\; x \in \mathbb{R}^N \right\}^1, \\ u_{k+1} = u_k + \gamma\big(Px_{k+1} - b\big), \quad \text{with} \quad \gamma > 0. \end{cases}$$

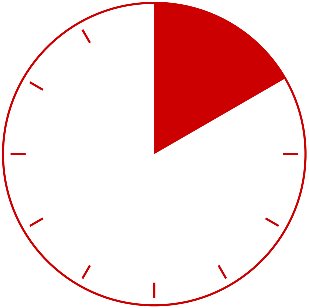
Proposition

If f is $\mu-$strongly convex, the Lagrangian Algorithm is convergent for $\gamma < 2\mu$, since it is equivalent to

$$\begin{cases} x_{k+1} = \nabla f^*(-P^T u_k), \\ u_{k+1} = u_k - \gamma\big(\nabla f^*(u_k) - b\big). \end{cases}$$

[1] Assuming this has a solution. When is it the case?

Break



The Method of Multipliers (Augmented Lagrangian)

We generate a sequence (x_k, u_k) by iterating

$$\begin{cases} x_{k+1} \in \operatorname{Argmin}\left\{ f(x) + u_k^T(Px - b) + \frac{\gamma}{2}\|Px - b\|^2 \ : \ x \in \mathbb{R}^N \right\}^2, \\ u_{k+1} = u_k + \gamma\big(Px_{k+1} - b\big), \quad \text{with} \quad \gamma > 0. \end{cases}$$

[2] Same as before

The Method of Multipliers (Augmented Lagrangian)

We generate a sequence (x_k, u_k) by iterating

$$\begin{cases} x_{k+1} \in \operatorname{Argmin} \left\{ f(x) + u_k^T(Px - b) + \frac{\gamma}{2}\|Px - b\|^2 \ : \ x \in \mathbb{R}^N \right\}^2, \\ u_{k+1} = u_k + \gamma\big(Px_{k+1} - b\big), \quad \text{with} \quad \gamma > 0. \end{cases}$$

Proposition

If f is convex and $\gamma > 0$, the method of multipliers is convergent, since it is equivalent to

$$\begin{cases} x_{k+1} = \partial f^*(-P^T u_k), \\ u_{k+1} = \operatorname{prox}_{\gamma d}(u_k), \quad \textit{where} \quad d(u) = f^*(-P^T u) + b \cdot u. \end{cases}$$

[2] Same as before

Variants

Are they convergent?

Proximal version

$$\begin{cases} x_{k+1} & = & \mathrm{Argmin}\left\{ f(x) + u_k^T(Px-b) + \dfrac{1}{2\gamma}\|x-x_k\|^2 \right\}, \\ u_{k+1} & = & u_k + \gamma(Px_{k+1}-b). \end{cases}$$

Variants

Are they convergent?

Proximal version

$$\begin{cases} x_{k+1} &=& \operatorname{Argmin}\left\{ f(x) + u_k^T(Px-b) + \dfrac{1}{2\gamma}\|x-x_k\|^2 \right\}, \\ u_{k+1} &=& u_k + \gamma(Px_{k+1}-b). \end{cases}$$

Predictor-corrector method

$$\begin{cases} p_{k+1} &=& u_k + \gamma(Px_k - b) \\ x_{k+1} &=& \operatorname{Argmin}\left\{ f(x) + p_{k+1}^T(Px-b) + \dfrac{1}{2\gamma}\|x-x_k\|^2 \right\} \\ u_{k+1} &=& u_k + \gamma(Px_{k+1}-b), \end{cases}$$

Summary

Exercise

1. Make a 4×3 table, with the methods above on the first column.

   On the second column, fill in the conditions on f required for well-posedness (for the iterations to be well defined).

   Finally, fill in the third column with the hypotheses you consider likely to ensure convergence. Explain the intuition behind your guess, outline a possible proof, or (even better) prove your claim.

2. Compare the methods (pros and cons) in terms of your findings.

Structured problems

Consider the structured problem

$$\min\left\{f(x)+g(y):(x,y)\in\mathbb{R}^{N_1\times N_2},\ Ax+By=c\right\}.$$

Structured problems

Consider the structured problem

$$\min \left\{ f(x) + g(y) : (x,y) \in \mathbb{R}^{N_1 \times N_2},\ Ax + By = c \right\}.$$

Questions

- What does the Lagrangian Algorithm give in this case?

Structured problems

Consider the structured problem

$$\min\left\{f(x)+g(y):(x,y)\in\mathbb{R}^{N_1\times N_2},\ Ax+By=c\right\}.$$

Questions

- What does the Lagrangian Algorithm give in this case?
- And the Method of Multipliers?

Structured problems

Consider the structured problem

$$\min \left\{ f(x) + g(y) : (x,y) \in \mathbb{R}^{N_1 \times N_2},\ Ax + By = c \right\}.$$

Questions

- What does the Lagrangian Algorithm give in this case?
- And the Method of Multipliers?
- Do you see any inconveniences?

Structured problems

Consider the structured problem

$$\min \left\{ f(x) + g(y) : (x, y) \in \mathbb{R}^{N_1 \times N_2},\ Ax + By = c \right\}.$$

Questions

- What does the Lagrangian Algorithm give in this case?
- And the Method of Multipliers?
- Do you see any inconveniences?
- What could we do instead?

Two proposals to solve $\min\{f(x)+g(y) : Ax+By=c\}$

When are they well-posed? Convergent? Implementable?

The Semi-Augmented Lagrangian Method is given by

$$\left\{\begin{array}{lcl} x_{k+1} & = & \operatorname{Argmin}\{f(x)+u_k\cdot Ax\} \\ y_{k+1} & = & \operatorname{Argmin}\{g(y)+u_k\cdot By+\frac{\gamma}{2}\|Ax_{k+1}+By-c\|^2\} \\ u_{k+1} & = & u_k+\gamma(Ax_{k+1}+By_{k+1}-c). \end{array}\right.$$

The Alternating Directions Method of Multipliers (ADMM) is

$$\left\{\begin{array}{lcl} x_{k+1} & = & \operatorname{Argmin}\{f(x)+u_k\cdot Ax+\frac{\gamma}{2}\|Ax+By_k-c\|^2\} \\ y_{k+1} & = & \operatorname{Argmin}\{g(y)+u_k\cdot By+\frac{\gamma}{2}\|Ax_{k+1}+By-c\|^2\} \\ u_{k+1} & = & u_k+\gamma(Ax_{k+1}+By_{k+1}-c). \end{array}\right.$$

Presentations

Proposed topics

1. Trust-region methods
2. Stochastic gradient
3. The Simplex Method
4. Interior point methods
5. Sequential Quadratic Programming

A standard description of these topics can be found, for instance, in Nocedal & Wright, Numerical Optimization, Springer, 2006.