Iterative Algorithms in Optimization, Variational Analysis and Fixed Point Theory

Unit 06: Equilibrium.

university of groningen

Zero Sum Games, Payoff Matrix

A (finite) game consists in

- A finite set $\mathcal{I}$ of players. There are $J = |\mathcal{I}|$ of them.

Zero Sum Games, Payoff Matrix

A (finite) game consists in

- A finite set $\mathcal{I}$ of players. There are $J = |\mathcal{I}|$ of them.
- Finite sets $S_1, \ldots, S_J$ of strategies for each player, which form a set $S = \prod S_i$ of strategy profiles.

Zero Sum Games, Payoff Matrix

A (finite) game consists in

- A finite set $\mathcal{I}$ of players. There are $J = |\mathcal{I}|$ of them.
- Finite sets $S_1, \ldots, S_J$ of strategies for each player, which form a set $S = \prod S_i$ of strategy profiles.
- Payoff functions $C_1, \ldots, C_J : S \to \mathbb{R}$ for each player.

Zero Sum Games, Payoff Matrix

A (finite) game consists in

- A finite set $\mathcal{I}$ of players. There are $J = |\mathcal{I}|$ of them.
- Finite sets $S_1, \ldots, S_J$ of strategies for each player, which form a set $S = \prod S_i$ of strategy profiles.
- Payoff functions $C_1, \ldots, C_J : S \to \mathbb{R}$ for each player.

A zero-sum game is one where $\sum C_i \equiv 0$.

Zero Sum Games, Payoff Matrix

A (finite) game consists in

- A finite set $\mathcal{I}$ of players. There are $J = |\mathcal{I}|$ of them.
- Finite sets $S_1, \ldots, S_J$ of strategies for each player, which form a set $S = \prod S_i$ of strategy profiles.
- Payoff functions $C_1, \ldots, C_J : S \to \mathbb{R}$ for each player.

A zero-sum game is one where $\sum C_i \equiv 0$.

In a two-player zero-sum game, all the payoffs can be arranged in one payoff matrix.

Pareto Efficiency and Nash Equilibria

A vector $v^* \in \mathbb{R}^J$ is Pareto-efficient or Pareto-optimal in $V \subset \mathbb{R}^J$ if there is no other $v \in V$ that dominates v^*, which means that

$$v_i \geq v_i^* \quad \text{for all} \quad i \in J \qquad \text{and} \qquad v_j > v_j^* \quad \text{for some} \quad j \in J.$$

Pareto Efficiency and Nash Equilibria

A vector $v^* \in \mathbb{R}^J$ is Pareto-efficient or Pareto-optimal in $V \subset \mathbb{R}^J$ if there is no other $v \in V$ that dominates v^*, which means that

$$v_i \geq v_i^* \quad \text{for all} \quad i \in J \qquad \text{and} \qquad v_j > v_j^* \quad \text{for some} \quad j \in J.$$

A strategy profile s^* is a Nash equilibrium of a game if the vector

$$\big(C_1(s^*), \ldots, C_J(s^*)\big)$$

is Pareto-efficient.

Matching Coins

Each player can draw either a nickel or a quarter.

Version 1

If at least one player draws a nickel, P1 gets both coins. Otherwise, P2 gets them.

Matching Coins

Each player can draw either a nickel or a quarter.

Version 1

If at least one player draws a nickel, P1 gets both coins. Otherwise, P2 gets them.

Payoff matrix:

$$\begin{pmatrix} 5 & 25 \\ 5 & -25 \end{pmatrix}.$$

Matching Coins

Each player can draw either a nickel or a quarter.

Version 1

If at least one player draws a nickel, P1 gets both coins. Otherwise, P2 gets them.

Payoff matrix:

$$\begin{pmatrix} 5 & 25 \\ 5 & -25 \end{pmatrix}.$$

Nash equilibrium: both players draw a nickel.

Matching Coins

Each player can draw either a nickel or a quarter.

Version 2

If P1 draws a nickel, P1 gets 5 cents from P2. If P2 draws a nickel and P1 draws a quarter, P1 gets 25 cents. If both draw quarters, P2 gets 25 cents.

Matching Coins

Each player can draw either a nickel or a quarter.

Version 2

If P1 draws a nickel, P1 gets 5 cents from P2. If P2 draws a nickel and P1 draws a quarter, P1 gets 25 cents. If both draw quarters, P2 gets 25 cents.

Payoff matrix:

$$\begin{pmatrix} 5 & 5 \\ 25 & -25 \end{pmatrix}.$$

Matching Coins

Each player can draw either a nickel or a quarter.

Version 2

If P1 draws a nickel, P1 gets 5 cents from P2. If P2 draws a nickel and P1 draws a quarter, P1 gets 25 cents. If both draw quarters, P2 gets 25 cents.

Payoff matrix:

$$\begin{pmatrix} 5 & 5 \\ 25 & -25 \end{pmatrix}.$$

Nash equilibrium: P1 draws a nickel, P2 a quarter.

Matching Coins

Each player can draw either a nickel or a quarter.

Version 3

If both players play the same coin, P1 gets them. Otherwise, P2 does.

Matching Coins

Each player can draw either a nickel or a quarter.

Version 3

If both players play the same coin, P1 gets them. Otherwise, P2 does.

Payoff matrix:

$$\begin{pmatrix} 5 & -25 \\ -5 & 25 \end{pmatrix}.$$

Matching Coins

Each player can draw either a nickel or a quarter.

Version 3

If both players play the same coin, P1 gets them. Otherwise, P2 does.

Payoff matrix:

$$\begin{pmatrix} 5 & -25 \\ -5 & 25 \end{pmatrix}.$$

Nash equilibrium: no (pure) Nash equilibrium.

Mixed strategies and the Minimax Problem

Players now use their strategies with a certain probability. A mixed strategy is a vector of probabilities over the set of strategies.

Mixed strategies and the Minimax Problem

Players now use their strategies with a certain probability. A mixed strategy is a vector of probabilities over the set of strategies.

Let A be the payoff matrix, while W and Z are sets of mixed strategies for Players 1 and 2, respectively.

Mixed strategies and the Minimax Problem

Players now use their strategies with a certain probability. A mixed strategy is a vector of probabilities over the set of strategies.

Let A be the payoff matrix, while W and Z are sets of mixed strategies for Players 1 and 2, respectively.

Player 1 tries to maximize

$$w^T A z,$$

by choosing $w \in W$, and Player 2 tries to minimize it, by choosing $z \in Z$.

Mixed strategies and the Minimax Problem

Von Neumann's Minimax Theorem

Let $W \subset \mathbb{R}^N$ and $Z \subset \mathbb{R}^M$ be convex and compact, and let $f : W \times Z \to \mathbb{R}$ be continuous and concave-convex. Then,

$$\min_{z \in Z} \left[\max_{w \in W} f(w, z) \right] = \max_{w \in W} \left[\min_{z \in Z} f(w, z) \right].$$

Mixed strategies and the Minimax Problem

Von Neumann's Minimax Theorem

Let $W \subset \mathbb{R}^N$ and $Z \subset \mathbb{R}^M$ be convex and compact, and let $f : W \times Z \to \mathbb{R}$ be continuous and concave-convex. Then,

$$\min_{z \in Z} \left[\max_{w \in W} f(w,z) \right] = \max_{w \in W} \left[\min_{z \in Z} f(w,z) \right].$$

Exercise

Show that every zero-sum game has a Nash equilibrium in mixed strategies. To this end, take $f(w,z) = w^T A z$

Matching Coins

Each player can draw either a nickel or a quarter.

Version 3

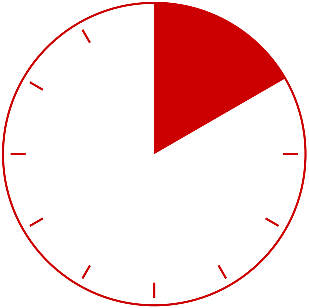
If both players play the same coin, P1 gets them. Otherwise, P2 does.

Payoff matrix:

$$\begin{pmatrix} 5 & -25 \\ -5 & 25 \end{pmatrix}.$$

Let us find the mixed Nash equilibrium!

Break



Saddle Points

A saddle point of $f : W \times Z \to \mathbb{R}$ is a point $(\hat{w}, \hat{z}) \in W \times Z$ such that

$$f(w, \hat{z}) \leq f(\hat{w}, \hat{z}) \leq f(\hat{w}, z)$$

for all $w \in W$ and $z \in Z$.

Saddle Points

A saddle point of $f : W \times Z \to \mathbb{R}$ is a point $(\hat{w}, \hat{z}) \in W \times Z$ such that

$$f(w, \hat{z}) \leq f(\hat{w}, \hat{z}) \leq f(\hat{w}, z)$$

for all $w \in W$ and $z \in Z$.

Example

In zero-sum games, the Nash equilibria in mixed strategies are the saddle points of $w^T Az$, with $w \in \Delta(S_1)$ and $z \in \Delta(S_2)$.

Optimization with constraints

Consider the problem

$$\min \left\{ f(x) \ : \ g_i(x) \leq 0, \ i = 1, \ldots, I; \ h_j(x) = 0, \ j = 1, \ldots J \right\}.$$

Optimization with constraints

Consider the problem

$$\min\left\{ f(x) \ : \ g_i(x) \leq 0, \ i = 1, \ldots, I; \ h_j(x) = 0, \ j = 1, \ldots J \right\}.$$

Optimality conditions

If $\hat{x}$ is a solution and $(\hat{\mu}, \hat{\lambda})$ are Lagrange multipliers, $(\hat{\mu}, \hat{\lambda}, \hat{x})$ is a saddle point of the Lagrangian

$$L(x, \mu, \lambda) = f(x) + \sum_i \mu_i g_i(x) + \sum_j \lambda_j h_j(x),$$

for $x \in \mathbb{R}^N$, $\mu \in \mathbb{R}_+^M$, $\lambda \in \mathbb{R}^L$.

Characterizing saddle points

A vector $(\hat{x}, \hat{y})$ is a saddle point of a convex-concave function f if, and only if

$$0 \in A \begin{pmatrix} x \\ y \end{pmatrix}.$$

where $A : X \times Y \rightrightarrows X \times Y$ is the saddle operator, defined by

$$A \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \partial_x f(x,y) \\ -\partial_y f(x,y) \end{pmatrix}.$$

Characterizing saddle points

A vector $(\hat{x}, \hat{y})$ is a saddle point of a convex-concave function f if, and only if

$$0 \in A \begin{pmatrix} x \\ y \end{pmatrix}.$$

where $A : X \times Y \rightrightarrows X \times Y$ is the saddle operator, defined by

$$A \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \partial_x f(x,y) \\ -\partial_y f(x,y) \end{pmatrix}.$$

Remark

The operator A is monotone. Under (mild) continuity assumptions, the operator $I + A$ is surjective and non-contracting.

Forward-backward algorithm

Let us recall the proximal-gradient algorithm:

$$x_{k+1} = (I + \gamma\partial g)^{-1}(x_k - \gamma\nabla f(x_k)).$$

Forward-backward algorithm

Let us recall the proximal-gradient algorithm:

$$x_{k+1} = (I + \gamma \partial g)^{-1}(x_k - \gamma \nabla f(x_k)).$$

We can extend this to monotone operators:

$$x_{k+1} = (I + \gamma A)^{-1}(x_k - \gamma B x_k),$$

whenever A is maximally monotone and B is cocoercive.

Iterative Algorithms in Optimization, Variational Analysis and Fixed Point Theory

Unit 06: Equilibrium.

university of groningen

Monotone operators

An operator $A : \mathbb{R}^N \rightrightarrows \mathbb{R}^N$ is monotone if

$$(x^* - y^*) \cdot (x - y) \geq 0$$

for all $x, y \in \mathbb{R}^N$, $x^* \in Ax$ and $y^* \in Ay$.

Monotone operators

An operator $A : \mathbb{R}^N \rightrightarrows \mathbb{R}^N$ is monotone if

$$(x^* - y^*) \cdot (x - y) \geq 0$$

for all $x, y \in \mathbb{R}^N$, $x^* \in Ax$ and $y^* \in Ay$.

Proposition

If A is monotone, then $I + A$ is non-contracting, which means that $\|(x + x^*) - (y + y^*)\| \geq \|x - y\|$ for all $x, y \in \mathbb{R}^N$, $x^* \in Ax$ and $y^* \in Ay$.

Monotone operators

An operator $A : \mathbb{R}^N \rightrightarrows \mathbb{R}^N$ is monotone if

$$(x^* - y^*) \cdot (x - y) \geq 0$$

for all $x, y \in \mathbb{R}^N$, $x^* \in Ax$ and $y^* \in Ay$.

Proposition

If A is monotone, then $I + A$ is non-contracting, which means that $\|(x + x^*) - (y + y^*)\| \geq \|x - y\|$ for all $x, y \in \mathbb{R}^N$, $x^* \in Ax$ and $y^* \in Ay$.

Examples

Subdifferentials and Saddle operators (see slide 13) are monotone.

Monotone operators

An operator $A : \mathbb{R}^N \rightrightarrows \mathbb{R}^N$ is monotone if

$$(x^* - y^*) \cdot (x - y) \geq 0$$

for all $x, y \in \mathbb{R}^N$, $x^* \in Ax$ and $y^* \in Ay$.

Proposition

If A is monotone, then $I + A$ is non-contracting, which means that $\|(x + x^*) - (y + y^*)\| \geq \|x - y\|$ for all $x, y \in \mathbb{R}^N$, $x^* \in Ax$ and $y^* \in Ay$.

Examples

Subdifferentials and Saddle operators (see slide 13) are monotone. If T is nonexpansive, $I - T$ is monotone.

Monotone operators

An operator $A : \mathbb{R}^N \rightrightarrows \mathbb{R}^N$ is monotone if

$$(x^* - y^*) \cdot (x - y) \geq 0$$

for all $x, y \in \mathbb{R}^N$, $x^* \in Ax$ and $y^* \in Ay$.

Proposition

If A is monotone, then $I + A$ is non-contracting, which means that $\|(x + x^*) - (y + y^*)\| \geq \|x - y\|$ for all $x, y \in \mathbb{R}^N$, $x^* \in Ax$ and $y^* \in Ay$.

Examples

Subdifferentials and Saddle operators (see slide 13) are monotone. If T is nonexpansive, $I - T$ is monotone. Sums, positive multiples and inverses of monotone operators are monotone.

Maximality

A maximally monotone operator is a monotone operator A for which $I + A$ is surjective.

Maximality

A maximally monotone operator is a monotone operator A for which $I + A$ is surjective.

Proposition

If A is maximally monotone, then

1. There is no other monotone operator that extends A.
2. λA is maximally monotone for every $\lambda > 0$.
3. A^{-1} is maximally monotone.

Maximality

A maximally monotone operator is a monotone operator A for which $I + A$ is surjective.

Proposition

If A is maximally monotone, then

1. There is no other monotone operator that extends A.
2. λA is maximally monotone for every $\lambda > 0$.
3. A^{-1} is maximally monotone.

The sum of two maximally monotone operators may not be maximal.

Resolvents

The resolvent of a maximally monotone operator A is the nonexpansive function $J_A : \mathbb{R}^N \to \mathbb{R}^N$, defined by

$$J_A x = (I + A)^{-1} x.$$

Resolvents

The resolvent of a maximally monotone operator A is the nonexpansive function $J_A : \mathbb{R}^N \to \mathbb{R}^N$, defined by

$$J_A x = (I + A)^{-1} x.$$

Example

If $f : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is closed and convex, then ∂f is maximally monotone, and

$$J_{\partial f} = \operatorname{prox}_f .$$

Cocoercive operators

A function $B : \mathbb{R}^N \to \mathbb{R}^N$ is $\theta-$cocoercive if

$$(Bx - By) \cdot (x - y) \geq \theta \|Bx - By\|^2$$

for all $x, y \in \mathbb{R}^N$.

Cocoercive operators

A function $B : \mathbb{R}^N \to \mathbb{R}^N$ is $\theta-$cocoercive if

$$(Bx - By) \cdot (x - y) \geq \theta \|Bx - By\|^2$$

for all $x, y \in \mathbb{R}^N$. If B is $\theta-$cocoercive, it is $\frac{1}{\theta}-$Lipschitz.

Cocoercive operators

A function $B : \mathbb{R}^N \to \mathbb{R}^N$ is $\theta-$cocoercive if

$$(Bx - By) \cdot (x - y) \geq \theta \|Bx - By\|^2$$

for all $x, y \in \mathbb{R}^N$. If B is $\theta-$cocoercive, it is $\frac{1}{\theta}-$Lipschitz.

Examples

If $f : \mathbb{R}^N \to \mathbb{R}$ is $L-$smooth and convex, then ∇f is $\frac{1}{L}-$cocoercive.
If T is nonexpansive, $I - T$ is $\frac{1}{2}-$cocoercive.

The forward-backward method

Let $A : \mathbb{R}^N \rightrightarrows \mathbb{R}^N$ be maximally monotone and let $B : \mathbb{R}^N \to \mathbb{R}^N$ be cocoercive.

The forward-backward method

Let $A : \mathbb{R}^N \rightrightarrows \mathbb{R}^N$ be maximally monotone and let $B : \mathbb{R}^N \to \mathbb{R}^N$ be cocoercive.

The forward-backward algorithm, defined by

$$X_{k+1} = J_{\gamma A}(x_k - \gamma B x_k),$$

allows us to approximate a zero of $A + B$, which happens to be maximal.

The forward-backward method

Let $A : \mathbb{R}^N \rightrightarrows \mathbb{R}^N$ be maximally monotone and let $B : \mathbb{R}^N \to \mathbb{R}^N$ be cocoercive.

The forward-backward algorithm, defined by

$$X_{k+1} = J_{\gamma A}(x_k - \gamma B x_k),$$

allows us to approximate a zero of $A + B$, which happens to be maximal.

Example

If $f : \mathbb{R}^N \to \mathbb{R}$ is smooth and convex, and $g : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is closed and convex, the forward-backward algorithm is the proximal-gradient method.

Convergence

Theorem

Let $A : \mathbb{R}^N \rightrightarrows \mathbb{R}^N$ be maximally monotone and let $B : \mathbb{R}^N \to \mathbb{R}^N$ be $\theta-$cocoercive. Assume that $S := \mathrm{Zer}(A + B) \neq \emptyset$, and that $\gamma \in (0, 2\theta)$. Every sequence generated by the forward-backward method converges to a point in S.

Convergence

Theorem

Let $A : \mathbb{R}^N \rightrightarrows \mathbb{R}^N$ be maximally monotone and let $B : \mathbb{R}^N \to \mathbb{R}^N$ be $\theta-$cocoercive. Assume that $S := \text{Zer}(A + B) \neq \emptyset$, and that $\gamma \in (0, 2\theta)$. Every sequence generated by the forward-backward method converges to a point in S.

Strategy

Show that there exist a nonexpansive function $T : \mathbb{R}^N \to \mathbb{R}^N$ and a number $\lambda \in (0, 1)$ such that $\text{Fix}(T) = \text{Zer}(A + B)$ and

$$J_{\gamma A} \circ (I - \gamma B) = \lambda I + (1 - \lambda) T.$$

Conclude from what we know about KM iterations.

Iterative Algorithms in Optimization, Variational Analysis and Fixed Point Theory

Unit 06: Equilibrium.

university of groningen

Convergence

Theorem

Let $A : \mathbb{R}^N \rightrightarrows \mathbb{R}^N$ be maximally monotone and let $B : \mathbb{R}^N \to \mathbb{R}^N$ be $\theta-$cocoercive. Assume that $S := \mathrm{Zer}(A + B) \neq \emptyset$, and that $\gamma \in (0, 2\theta)$. Every sequence generated by the forward-backward method converges to a point in S.

Convergence

Theorem

Let $A : \mathbb{R}^N \rightrightarrows \mathbb{R}^N$ be maximally monotone and let $B : \mathbb{R}^N \to \mathbb{R}^N$ be $\theta-$cocoercive. Assume that $S := \mathrm{Zer}(A + B) \neq \emptyset$, and that $\gamma \in (0, 2\theta)$. Every sequence generated by the forward-backward method converges to a point in S.

Strategy

Show that there exist a nonexpansive function $T : \mathbb{R}^N \to \mathbb{R}^N$ and a number $\lambda \in (0, 1)$ such that $\mathrm{Fix}(T) = \mathrm{Zer}(A + B)$ and

$$J_{\gamma A} \circ (I - \gamma B) = \lambda I + (1 - \lambda) T.$$

Conclude from what we know about KM iterations.

Variations

A maximally monotone, B monotone and Lipschitz

Variations

A maximally monotone, B monotone and Lipschitz

Korpelevich, 1976

$$\left\{\begin{array}{rcl} y_k & = & \mathrm{Proj}_C(x_k - \gamma B x_k) \\ x_{k+1} & = & \mathrm{Proj}_C(x_k - \gamma B y_k). \end{array}\right.$$

Variations

A maximally monotone, B monotone and Lipschitz

Korpelevich, 1976

$$\left\{\begin{array}{rcl} y_k & = & \mathrm{Proj}_C(x_k - \gamma B x_k) \\ x_{k+1} & = & \mathrm{Proj}_C(x_k - \gamma B y_k). \end{array}\right.$$

Tseng, 2000

$$\left\{\begin{array}{rcl} y_k & = & J_{\gamma A}(x_k - \gamma B x_k) \\ x_{k+1} & = & x_k - \gamma\big(B x_k - B y_k\big). \end{array}\right.$$

Variations

A maximally monotone, *B* monotone and Lipschitz

Korpelevich, 1976

$$\left\{\begin{array}{rcl} y_k & = & \mathrm{Proj}_C(x_k - \gamma B x_k) \\ x_{k+1} & = & \mathrm{Proj}_C(x_k - \gamma B y_k). \end{array}\right.$$

Tseng, 2000

$$\left\{\begin{array}{rcl} y_k & = & J_{\gamma A}(x_k - \gamma B x_k) \\ x_{k+1} & = & x_k - \gamma\big(B x_k - B y_k\big). \end{array}\right.$$

Malytski-Tam, 2020

$$x_{k+1} = J_{\gamma A}\big(x_k - 2\gamma B x_k + \gamma B x_{k-1}\big).$$

3-operator splitting

A, B maximally monotone, C cocoercive

Davis-Yin, 2017

$$\begin{cases} y_k &= J_{\gamma B} x_k \\ z_k &= J_{\gamma A}(2y_k - x_k - \gamma C y_k) \\ x_{k+1} &= x_k + \lambda_k (z_k - y_k). \end{cases}$$

3-operator splitting

A, B maximally monotone, C cocoercive

Davis-Yin, 2017

$$\left\{ \begin{array}{rcl} y_k & = & J_{\gamma B} x_k \\ z_k & = & J_{\gamma A}(2y_k - x_k - \gamma C y_k) \\ x_{k+1} & = & x_k + \lambda_k(z_k - y_k). \end{array} \right.$$

Case $B = 0$: Forward-backward.

3-operator splitting

A, B maximally monotone, C cocoercive

Davis-Yin, 2017

$$\left\{ \begin{array}{rcl} y_k & = & J_{\gamma B} x_k \\ z_k & = & J_{\gamma A}(2y_k - x_k - \gamma C y_k) \\ x_{k+1} & = & x_k + \lambda_k (z_k - y_k). \end{array} \right.$$

Case $B = 0$: Forward-backward.

Case $C = 0$: Douglas-Rachford, 1956

- Originally introduced to solve PDE's.

Recall Primal-Dual in optimization

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

Recall Primal-Dual in optimization

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

Implementation trick: Moreau's Identity

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