Introduction to the Augmented Lagrangian Method

Pranav Reddy

University of California, San Diego

6 March 2024

Overview

Goals

1. Introduce the augmented Lagrangian method (ALM)

Goals

1. Introduce the augmented Lagrangian method (ALM)
2. Understand optimality conditions for differentiable optimization

Goals

1. Introduce the augmented Lagrangian method (ALM)
2. Understand optimality conditions for differentiable optimization
3. Prove properties of the ALM

Goals

1. Introduce the augmented Lagrangian method (ALM)
2. Understand optimality conditions for differentiable optimization
3. Prove properties of the ALM
4. Understand limitations of ALM iterations

Goals

1. Introduce the augmented Lagrangian method (ALM)
2. Understand optimality conditions for differentiable optimization
3. Prove properties of the ALM
4. Understand limitations of ALM iterations
5. Discuss future directions

Motivation

Let us consider the primal problem (P)

$$\begin{aligned} \min_{x \in \mathcal{X}} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0. \quad i = 1, \ldots, m \end{aligned} \tag{P}$$

where $f, g_1, \ldots, g_m$ are convex functions and $\mathcal{X}$ is a convex set.

Motivation

Let us consider the primal problem (P)

$$\min_{x \in \mathcal{X}} \quad f(x) \tag{P}$$
$$\text{s.t.} \quad g_i(x) \le 0. \quad i = 1, \ldots, m$$

where $f, g_1, \ldots, g_m$ are convex functions and $\mathcal{X}$ is a convex set.
The **Lagrangian** associated with (P) is $L\colon \mathbb{R}^n \times \mathbb{R}^m_+ \to \mathbb{R}$ defined by

$$L(x, y) = f(x) + \sum_{i=1}^{m} y_i g_i(x) \tag{L}$$

Motivation

Let us consider the primal problem (P)

$$\min_{x \in \mathcal{X}} \quad f(x) \qquad \text{(P)}$$
$$\text{s.t.} \quad g_i(x) \leq 0. \quad i = 1, \ldots, m$$

where $f, g_1, \ldots, g_m$ are convex functions and $\mathcal{X}$ is a convex set.
The **Lagrangian** associated with (P) is $L \colon \mathbb{R}^n \times \mathbb{R}^m_+ \to \mathbb{R}$ defined by

$$L(x, y) = f(x) + \sum_{i=1}^{m} y_i g_i(x) \qquad \text{(L)}$$

The **Lagrange dual function** is

$$g_0(y) = \inf_{x \in \mathcal{X}} L(x, y)$$

Duality Continued

This gives the dual problem (D)

$$d^\star = \max_{y \in \mathbb{R}^m_+} \quad g_0(y) \tag{D}$$

Duality Continued

This gives the dual problem (D)

$$d^\star = \max_{y \in \mathbb{R}^m_+} \quad g_0(y) \tag{D}$$

- Note that in general $p^\star \geq d^\star$. This is known as **weak duality**.

Duality Continued

This gives the dual problem (D)

$$d^\star = \max_{y \in \mathbb{R}^m_+} \quad g_0(y) \tag{D}$$

- Note that in general $p^\star \geq d^\star$. This is known as **weak duality**.
- If $p^\star = d^\star$, then we say that **strong duality** holds.

Duality Continued

This gives the dual problem (D)

$$d^\star = \max_{y \in \mathbb{R}^m_+} \quad g_0(y) \tag{D}$$

- Note that in general $p^\star \geq d^\star$. This is known as **weak duality**.
- If $p^\star = d^\star$, then we say that **strong duality** holds.
- A natural question is to ask when a point (x, y) is optimal for (P) and (D)

KKT Conditions

Theorem (KKT Conditions (Necessity))

Let $(x^\star, y^\star)$ *be an optimal pair for* (P) *and* (D) *respectively. Assume strong duality holds, so* $f(x^\star) = p^\star = d^\star = g_0(y^\star)$*. Then, the following are true:*

1. $g_i(x^\star) \leq 0$ *(primal feasibility)*
2. $y_i^\star \geq 0$ *(dual feasibility)*
3. $y_i^\star g_i(x^\star) = 0$ *(strict complementarity)*
4. $\nabla_x f(x^\star) + \sum_{i=1}^{m} y_i^\star \nabla_x g_i(x^\star) = 0$ *(first-order condition)*

Proof of KKT Conditions

Proof.

Let $(x^\star, y^\star)$ be an optimal pair for (P) and (D) respectively. Then primal and dual feasibility follow immediately.

Proof of KKT Conditions

Proof.

Let $(x^\star, y^\star)$ be an optimal pair for (P) and (D) respectively. Then primal and dual feasibility follow immediately. For the other constraints

□

Proof of KKT Conditions

Proof.

Let $(x^\star, y^\star)$ be an optimal pair for (P) and (D) respectively. Then primal and dual feasibility follow immediately. For the other constraints

$$f(x^\star) = g_0(y^\star)$$

□

Proof of KKT Conditions

Proof.

Let $(x^\star, y^\star)$ be an optimal pair for (P) and (D) respectively. Then primal and dual feasibility follow immediately. For the other constraints

$$
\begin{aligned}
f(x^\star) &= g_0(y^\star) \\
&= \inf_{x \in \mathcal{X}} L(x, y^\star)
\end{aligned}
$$

□

Proof of KKT Conditions

Proof.

Let $(x^\star, y^\star)$ be an optimal pair for (P) and (D) respectively. Then primal and dual feasibility follow immediately. For the other constraints

$$
\begin{aligned}
f(x^\star) &= g_0(y^\star) \\
&= \inf_{x \in \mathcal{X}} L(x, y^\star) \\
&= \inf_{x \in \mathcal{X}} f(x) + \sum_{i=1}^{m} y_i^\star g_i(x)
\end{aligned}
$$

□

Proof of KKT Conditions

Proof.

Let $(x^\star, y^\star)$ be an optimal pair for (P) and (D) respectively. Then primal and dual feasibility follow immediately. For the other constraints

$$\begin{aligned} f(x^\star) &= g_0(y^\star) \\ &= \inf_{x \in \mathcal{X}} L(x, y^\star) \\ &= \inf_{x \in \mathcal{X}} f(x) + \sum_{i=1}^{m} y_i^\star g_i(x) \\ &\leq f(x^\star) + \sum_{i=1}^{m} y_i^\star g_i(x^\star) \end{aligned}$$

□

Proof of KKT Conditions

Proof.

Let $(x^\star, y^\star)$ be an optimal pair for (P) and (D) respectively. Then primal and dual feasibility follow immediately. For the other constraints

$$
\begin{aligned}
f(x^\star) &= g_0(y^\star) \\
&= \inf_{x \in \mathcal{X}} L(x, y^\star) \\
&= \inf_{x \in \mathcal{X}} f(x) + \sum_{i=1}^{m} y_i^\star g_i(x) \\
&\leq f(x^\star) + \sum_{i=1}^{m} y_i^\star g_i(x^\star) \\
&\leq f(x^\star)
\end{aligned}
$$

□

Proof of KKT Condition Cont.

This implies the KKT conditions immediately

- $x^\star$ minimizes $L(x, y^\star)$.

Proof of KKT Condition Cont.

This implies the KKT conditions immediately

- $x^\star$ minimizes $L(x, y^\star)$.
- Therefore, $\sum_{i=1}^{m} y_i^\star g_i(x^\star) = 0$, and therefore $y_i^\star g_i(x^\star) = 0$.

Proof of KKT Condition Cont.

This implies the KKT conditions immediately

- $x^\star$ minimizes $L(x, y^\star)$.
- Therefore, $\sum_{i=1}^m y_i^\star g_i(x^\star) = 0$, and therefore $y_i^\star g_i(x^\star) = 0$.
- Since $x^\star$ minimizes $L(x, y^\star)$, $\nabla_x L(x^\star, y^\star) = 0$.

KKT Conditions Cont.

Theorem (KKT Conditions (Sufficient))

Assume strong duality holds, so $p^\star = d^\star$. *Let* $(x^\star, y^\star)$ *be such that the following are true.*

1. $g_i(x^\star) \leq 0$ *(primal feasibility)*
2. $y_i^\star \geq 0$ *(dual feasibility)*
3. $y_i^\star g_i(x^\star) = 0$ *(strict complementarity)*
4. $\nabla_x f(x^\star) + \sum_{i=1}^m y_i^\star \nabla_x g_i(x^\star) = 0$ *(first-order condition)*

Then $(x^\star, y^\star)$ *are an optimal pair for* (P) *and* (D) *respectively.*

Proof of KKT Conditions

Proof.

Note that

$$g_0(y^\star) = \inf_{x \in \mathcal{X}} L(x, y^\star)$$

□

Proof of KKT Conditions

Proof.

Note that

$$
\begin{aligned}
g_0(y^\star) &= \inf_{x \in \mathcal{X}} L(x, y^\star) \\
&= \inf_{x \in \mathcal{X}} f(x) + \sum_{i=1}^{m} y_i^\star g_i(x)
\end{aligned}
$$

□

Proof of KKT Conditions

Proof.

Note that

$$\begin{aligned} g_0(y^\star) &= \inf_{x \in \mathcal{X}} L(x, y^\star) \\ &= \inf_{x \in \mathcal{X}} f(x) + \sum_{i=1}^{m} y_i^\star g_i(x) \\ &= f(x^\star) + \sum_{i=1}^{m} y_i^\star g_i(x^\star) \end{aligned}$$

□

Proof of KKT Conditions

Proof.

Note that

$$\begin{aligned} g_0(y^\star) &= \inf_{x \in \mathcal{X}} L(x, y^\star) \\ &= \inf_{x \in \mathcal{X}} f(x) + \sum_{i=1}^{m} y_i^\star g_i(x) \\ &= f(x^\star) + \sum_{i=1}^{m} y_i^\star g_i(x^\star) \\ &= f(x^\star) \end{aligned}$$

Note that we used the convexity of the problem here to assert that a stationary point is a minimizer. □

Application of the KKT Conditions

This suggests some ideas for finding minima: we can minimize the primal variable and maximize the dual variable to satisfy the KKT conditions:

$$
\begin{aligned}
x^{(k+1)} &= \operatorname*{argmin}_{x \in \mathcal{X}} f(x) + \sum_{i=1}^{m} y_i^{(k)} g_i(x) \\
y^{(k+1)} &= y^{(k)} + c^{(k)} \nabla_y \left(f(x^{(k+1)}) + \sum_{i=1}^{m} y_i^{(k)} g_i(x^{(k+1)}) \right) \\
&= y^{(k)} + c^{(k)} \begin{bmatrix} g_1(x^{(k+1)}) \\ \vdots \\ g_m(x^{(k+1)}) \end{bmatrix}
\end{aligned}
$$

Application of the KKT Conditions

$$
\begin{aligned}
x^{(k+1)} &= \operatorname*{argmin}_{x \in \mathcal{X}} f(x) + \sum_{i=1}^{m} y_i^{(k)} g_i(x) \\
y^{(k+1)} &= y^{(k)} + c^{(k)} \nabla_y \left(f(x^{(k+1)}) + \sum_{i=1}^{m} y_i^{(k)} g_i(x^{(k+1)}) \right) \\
&= y^{(k)} + c^{(k)} \begin{bmatrix} g_1(x^{(k+1)}) \\ \vdots \\ g_m(x^{(k+1)}) \end{bmatrix}
\end{aligned}
$$

However, this method is difficult to use in practice. Rockafellar [3] gives some reasons why this method is undesirable

Application of the KKT Conditions

$$x^{(k+1)} = \underset{x \in \mathcal{X}}{\operatorname{argmin}} f(x) + \sum_{i=1}^{m} y_i^{(k)} g_i(x)$$

$$y^{(k+1)} = y^{(k)} + c^{(k)} \nabla_y \left(f(x^{(k+1)}) + \sum_{i=1}^{m} y_i^{(k)} g_i(x^{(k+1)}) \right)$$

$$= y^{(k)} + c^{(k)} \begin{bmatrix} g_1(x^{(k+1)}) \\ \vdots \\ g_m(x^{(k+1)}) \end{bmatrix}$$

However, this method is difficult to use in practice. Rockafellar [3] gives some reasons why this method is undesirable

- It is hard to ensure that $y^{(k)} \geq 0$ for all iterations

Application of the KKT Conditions

$$x^{(k+1)} = \underset{x \in \mathcal{X}}{\operatorname{argmin}}\, f(x) + \sum_{i=1}^{m} y_i^{(k)} g_i(x)$$

$$\begin{aligned} y^{(k+1)} &= y^{(k)} + c^{(k)} \nabla_y \left(f(x^{(k+1)}) + \sum_{i=1}^{m} y_i^{(k)} g_i(x^{(k+1)}) \right) \\ &= y^{(k)} + c^{(k)} \begin{bmatrix} g_1(x^{(k+1)}) \\ \vdots \\ g_m(x^{(k+1)}) \end{bmatrix} \end{aligned}$$

However, this method is difficult to use in practice. Rockafellar [3] gives some reasons why this method is undesirable

- It is hard to ensure that $y^{(k)} \geq 0$ for all iterations
- $x^{(k+1)}$ may or may not exist at each iteration

Application of the KKT Conditions

$$x^{(k+1)} = \underset{x\in\mathcal{X}}{\operatorname{argmin}}\, f(x) + \sum_{i=1}^{m} y_i^{(k)} g_i(x)$$

$$y^{(k+1)} = y^{(k)} + c^{(k)} \nabla_y \left(f(x^{(k+1)}) + \sum_{i=1}^{m} y_i^{(k)} g_i(x^{(k+1)}) \right)$$

$$= y^{(k)} + c^{(k)} \begin{bmatrix} g_1(x^{(k+1)}) \\ \vdots \\ g_m(x^{(k+1)}) \end{bmatrix}$$

However, this method is difficult to use in practice. Rockafellar [3] gives some reasons why this method is undesirable

- It is hard to ensure that $y^{(k)} \geq 0$ for all iterations
- $x^{(k+1)}$ may or may not exist at each iteration
- If f is not strictly convex, $\{x^{(k)}\}$ may not converge to a minimizer

Penalty Functions

Let us consider the problem (P) again

$$
\begin{aligned}
p^\star = \min_{x\in\mathcal{X}} \quad & f(x) \\
\text{s.t.} \quad & g_i(x) \le 0. \quad i = 1, \ldots, m
\end{aligned}
\tag{(P)}
$$

Penalty Functions

Let us consider the problem (P) again

$$p^\star = \min_{x \in \mathcal{X}} \quad f(x) \qquad \text{((P))}$$
$$\text{s.t.} \quad g_i(x) \leq 0. \quad i = 1, \ldots, m$$

One approach to deal with the constraint is to move it into the cost, like so

$$\min_{x \in \mathcal{X}} \quad f(x) + \rho \sum_{i=1}^{m} \max\{0, g_i(x)\}$$

where ρ is a penalty parameter. Note that as $\rho \to \infty$ we recover the original problem.

Penalized Formulation

Frequently, we will also consider

$$\min_{x \in \mathcal{X}} \quad f(x) + \rho \sum_{i=1}^{m} \max\{0, g_i(x)\}^2$$

Penalized Formulation

Frequently, we will also consider

$$\min_{x \in \mathcal{X}} \quad f(x) + \rho \sum_{i=1}^{m} \max\{0, g_i(x)\}^2$$

- Both the linear and quadratic penalty formulations suffer from numerical issues as $\rho \to \infty$

Penalized Formulation

Frequently, we will also consider

$$\min_{x \in \mathcal{X}} \quad f(x) + \rho \sum_{i=1}^{m} \max\{0, g_i(x)\}^2$$

- Both the linear and quadratic penalty formulations suffer from numerical issues as $\rho \to \infty$
- Additionally, this is still not entirely an unconstrained problem

Penalized Formulation

Frequently, we will also consider

$$\min_{x \in \mathcal{X}} \quad f(x) + \rho \sum_{i=1}^{m} \max\{0, g_i(x)\}^2$$

- Both the linear and quadratic penalty formulations suffer from numerical issues as $\rho \to \infty$
- Additionally, this is still not entirely an unconstrained problem
- Moreover, the objective function may not be differentiable anymore

Penalized Formulation

Frequently, we will also consider

$$\min_{x \in \mathcal{X}} \quad f(x) + \rho \sum_{i=1}^{m} \max\{0, g_i(x)\}^2$$

- Both the linear and quadratic penalty formulations suffer from numerical issues as $\rho \to \infty$
- Additionally, this is still not entirely an unconstrained problem
- Moreover, the objective function may not be differentiable anymore
- To treat this issue, we will need a different formulation

The Augmented Lagrangian

From Rockafellar [3], we can consider the following generalization of the Lagrangian

Definition

The **augmented Lagrangian** of the problem (P) is $L_r \colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ defined as

$$L_r(x,y) \triangleq f(x) + \frac{1}{4r}\left(\sum_{i=1}^{m} \max\{0, y_i + 2r \cdot g_i(x)\}^2 - \|y\|_2^2\right) \quad \text{(AL)}$$

The Augmented Lagrangian

From Rockafellar [3], we can consider the following generalization of the Lagrangian

Definition

The **augmented Lagrangian** of the problem (P) is $L_r \colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ defined as

$$L_r(x, y) \triangleq f(x) + \frac{1}{4r}\left(\sum_{i=1}^{m} \max\{0, y_i + 2r \cdot g_i(x)\}^2 - \|y\|_2^2\right) \quad \text{(AL)}$$

The **augmented Lagrangian method** (**ALM**) of Rockafellar [2] is

$$\begin{aligned} x^{(k+1)} &= \operatorname*{argmin}_{x \in \mathcal{X}} L_{c^{(k)}}(x, y^{(k)}) \\ y^{(k+1)} &= y^{(k)} + c^{(k)} \nabla_y L_{c^{(k)}}(x^{(k+1)}, y^{(k)}) \end{aligned} \quad \text{(ALM)}$$

The Augmented Lagrangian

From Rockafellar [3], we can consider the following generalization of the Lagrangian

Definition

The **augmented Lagrangian** of the problem (P) is $L_r \colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ defined as

$$L_r(x, y) \triangleq f(x) + \frac{1}{4r} \left(\sum_{i=1}^{m} \max\{0, y_i + 2r \cdot g_i(x)\}^2 - \|y\|_2^2 \right) \quad \text{(AL)}$$

The **augmented Lagrangian method** (**ALM**) of Rockafellar [2] is

$$\begin{aligned} x^{(k+1)} &= \operatorname*{argmin}_{x \in \mathcal{X}} L_{c^{(k)}}(x, y^{(k)}) \\ y^{(k+1)} &= y^{(k)} + c^{(k)} \nabla_y L_{c^{(k)}}(x^{(k+1)}, y^{(k)}) \end{aligned} \quad \text{(ALM)}$$

In practice, we typically cannot solve the primal update $x^{(k+1)}$ update exactly, so we allow for some degree of inexactness.

Properties of the Augmented Lagrangian

Theorem

For any $r \geq 0$, we have

$$L_r(x, y) = \min\{F_r(x, u) + \langle u, y \rangle \mid u \in \mathbb{R}^m\} \quad x \in \mathcal{X}$$

where

$$F_r(x, u) = \begin{cases} f_0(x) + r\|u\|_2^2 & \text{if } u_i \geq g_i(x), \quad i = 1, \ldots, m \\ +\infty & \text{otherwise} \end{cases}$$

Moreover, $L_r(x, y)$ is convex in x and concave in y.

Proof

Proof.

The convexity and concavity follows if we can prove the equality. We prove slightly more general property instead. Consider

$$\begin{aligned} \min_{x \in \mathcal{X}, v \geq 0} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) + v_i = 0. \quad i = 1, \ldots, m \end{aligned}$$

Proof

Proof.

The convexity and concavity follows if we can prove the equality. We prove slightly more general property instead. Consider

$$\begin{aligned} \min_{x \in \mathcal{X}, v \geq 0} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) + v_i = 0. \quad i = 1, \ldots, m \end{aligned}$$

The corresponding penalty formulation for this problem is

$$L_r(x, v, y) = f(x) + r \sum_{i=1}^{m} (g_i(x) + v_i)^2 + \sum_{i=1}^{m} y_i (g_i(x) + v_i)$$

Proof

Proof.

The convexity and concavity follows if we can prove the equality. We prove slightly more general property instead. Consider

$$\begin{aligned} \min_{x \in \mathcal{X}, v \geq 0} \quad & f(x) \\ \text{s.t.} \quad & g_i(x) + v_i = 0. \quad i = 1, \ldots, m \end{aligned}$$

The corresponding penalty formulation for this problem is

$$L_r(x, v, y) = f(x) + r \sum_{i=1}^{m} (g_i(x) + v_i)^2 + \sum_{i=1}^{m} y_i (g_i(x) + v_i)$$

Minimizing this in v gives

$$v_i = \max\left\{0, g_i(x) - \frac{2z_i}{r}\right\}$$

Proof Cont.

Proof.

The corresponding penalty formulation for this problem is

$$L_r(x, v, y) = f(x) + r\sum_{i=1}^{m}(g_i(x) + v_i)^2 + \sum_{i=1}^{m} y_i(g_i(x) + v_i)$$

Minimizing this in v gives

$$v_i = \max\left\{0, -\frac{2z_i}{r} - g_i(x)\right\}$$

Plugging this in and rearranging gives (AL)

$$f(x) + \frac{1}{4r}\left(\sum_{i=1}^{m} \max\{0, y_i + 2r \cdot g_i(x)\}^2 - \|y\|_2^2\right) = L_r(x, y)$$

Choosing $v_i = z_i + g_i(x)$ completes the proof. □

Corresponding Dual

The dual function corresponding to the augmented Lagrangian (AL) is

$$g_r(y) = \inf_{x \in \mathcal{X}} L_r(x, y)$$

Corresponding Dual

The dual function corresponding to the augmented Lagrangian (AL) is

$$g_r(y) = \inf_{x \in \mathcal{X}} L_r(x, y)$$

The dual problem is then

$$d_r^\star = \max_{y \in \mathbb{R}^m} \quad g_r(x, y) \tag{$\mathrm{D_r}$}$$

Characterization of ALM Dual Problem

Theorem

For all $r > 0$, $g_r(y)$ is concave and satisfies

$$g_r(y) = \max_{z \in \mathbb{R}^m} \left\{ g_0(z) - \frac{1}{4r} \|z - y\|^2 \right\}$$

and the problem $(\mathrm{D_r})$ *has the same optimal solutions as* (D). *If g_0 is not $-\infty$ everywhere, then g_r is finite everywhere and $\mathcal{C}^1$.*
Additionally, if for a given y there exists x such that $g_r(y) = L_r(x, y)$, then

$$\frac{\partial g_r(y)}{\partial y_i} = \max \left\{ \frac{-y_i}{2r}, g_i(x) \right\}$$

Proof

Proof.

Let

$$p_r(u) \triangleq \min_{x \in \mathcal{X}} F_r(x, u)$$

$$q(u) \triangleq \frac{1}{2}\|u\|^2$$

Note that p_r is convex, since F_r is convex and by expanding the definition of $F_r(x, u)$, we can see that $p_r(u) = p_0(u) + 2rq(u)$.

Proof

Proof.

Let

$$p_r(u) \triangleq \min_{x \in \mathcal{X}} F_r(x, u)$$
$$q(u) \triangleq \frac{1}{2}\|u\|^2$$

Note that p_r is convex, since F_r is convex and by expanding the definition of $F_r(x, u)$, we can see that $p_r(u) = p_0(u) + 2rq(u)$. Then,

$$g_r(y) = \inf_{x \in \mathcal{X}} L_r(x, y)$$

□

Proof

Proof.

Let

$$p_r(u) \triangleq \min_{x \in \mathcal{X}} F_r(x, u)$$

$$q(u) \triangleq \frac{1}{2}\|u\|^2$$

Note that p_r is convex, since F_r is convex and by expanding the definition of $F_r(x, u)$, we can see that $p_r(u) = p_0(u) + 2rq(u)$.

Then,

$$\begin{aligned} g_r(y) &= \inf_{x \in \mathcal{X}} L_r(x, y) \\ &= \inf_{x \in \mathcal{X}} \min_{u \in \mathbb{R}^m} (F_r(x, u) + \langle u, y \rangle) \end{aligned}$$

□

Proof

Proof.

Let

$$p_r(u) \triangleq \min_{x \in \mathcal{X}} F_r(x, u)$$

$$q(u) \triangleq \frac{1}{2}\|u\|^2$$

Note that p_r is convex, since F_r is convex and by expanding the definition of $F_r(x, u)$, we can see that $p_r(u) = p_0(u) + 2rq(u)$.

Then,

$$\begin{aligned} g_r(y) &= \inf_{x \in \mathcal{X}} L_r(x, y) \\ &= \inf_{x \in \mathcal{X}} \min_{u \in \mathbb{R}^m} (F_r(x, u) + \langle u, y \rangle) \\ &= \inf_{u \in \mathbb{R}^m} (p_r(u) + \langle u, y \rangle) \end{aligned}$$

□

Proof

Proof.

Let

$$p_r(u) \triangleq \min_{x \in \mathcal{X}} F_r(x, u)$$

$$q(u) \triangleq \frac{1}{2}\|u\|^2$$

Note that p_r is convex, since F_r is convex and by expanding the definition of $F_r(x, u)$, we can see that $p_r(u) = p_0(u) + 2rq(u)$.

Then,

$$\begin{aligned} g_r(y) &= \inf_{x \in \mathcal{X}} L_r(x, y) \\ &= \inf_{x \in \mathcal{X}} \min_{u \in \mathbb{R}^m} (F_r(x, u) + \langle u, y \rangle) \\ &= \inf_{u \in \mathbb{R}^m} (p_r(u) + \langle u, y \rangle) \\ &= -p_r^{\star}(-y) \quad \textit{(convex conjugate)} \end{aligned}$$

□

Proof Cont.

Proof.

$$
\begin{aligned}
g_r(y) &= \inf_{x\in\mathcal{X}} L_r(x,y) \\
&= \inf_{x\in\mathcal{X}} \min_{u\in\mathbb{R}^m} \left(F_r(x,u) + \langle u, y\rangle\right) \\
&= \inf_{u\in\mathbb{R}^m} \left(p_r(u) + \langle u, y\rangle\right) \\
&= -p_r^\star(-y) \quad \textit{(convex conjugate)}
\end{aligned}
$$

Proof Cont.

Proof.

$$
\begin{aligned}
g_r(y) &= \inf_{x \in \mathcal{X}} L_r(x, y) \\
&= \inf_{x \in \mathcal{X}} \min_{u \in \mathbb{R}^m} \left(F_r(x, u) + \langle u, y \rangle\right) \\
&= \inf_{u \in \mathbb{R}^m} \left(p_r(u) + \langle u, y \rangle\right) \\
&= -p_r^\star(-y) \quad \textit{(convex conjugate)} \\
&= -(p_0 + 2rq)^\star(-y)
\end{aligned}
$$

Proof Cont.

Proof.

$$
\begin{aligned}
g_r(y) &= \inf_{x\in\mathcal{X}} L_r(x,y) \\
&= \inf_{x\in\mathcal{X}} \min_{u\in\mathbb{R}^m} \left(F_r(x,u) + \langle u, y\rangle\right) \\
&= \inf_{u\in\mathbb{R}^m} \left(p_r(u) + \langle u, y\rangle\right) \\
&= -p_r^\star(-y) \quad \textit{(convex conjugate)} \\
&= -(p_0 + 2rq)^\star(-y) \\
&= -(p_0^\star \,\square\, 2rq^\star)(-y) \quad \textit{(infimal convolution)}
\end{aligned}
$$

Proof Cont.

Proof.

$$
\begin{aligned}
g_r(y) &= \inf_{x\in\mathcal{X}} L_r(x,y) \\
&= \inf_{x\in\mathcal{X}} \min_{u\in\mathbb{R}^m} (F_r(x,u) + \langle u, y\rangle) \\
&= \inf_{u\in\mathbb{R}^m} (p_r(u) + \langle u, y\rangle) \\
&= -p_r^\star(-y) \quad \textit{(convex conjugate)} \\
&= -(p_0 + 2rq)^\star(-y) \\
&= -(p_0^\star \,\square\, 2rq^\star)(-y) \quad \textit{(infimal convolution)} \\
&= -\inf_{z\in\mathbb{R}^m} \left\{ p_0^\star(-z) + 2rq^\star\left(\frac{y-z}{2r}\right)\right\}
\end{aligned}
$$

Proof Cont.

Proof.

$$
\begin{aligned}
g_r(y) &= \inf_{x\in\mathcal{X}} L_r(x,y) \\
&= \inf_{x\in\mathcal{X}} \min_{u\in\mathbb{R}^m} \left(F_r(x,u) + \langle u, y\rangle\right) \\
&= \inf_{u\in\mathbb{R}^m} \left(p_r(u) + \langle u, y\rangle\right) \\
&= -p_r^\star(-y) \quad \textit{(convex conjugate)} \\
&= -(p_0 + 2rq)^\star(-y) \\
&= -(p_0^\star \,\square\, 2rq^\star)(-y) \quad \textit{(infimal convolution)} \\
&= -\inf_{z\in\mathbb{R}^m} \left\{ p_0^\star(-z) + 2rq^\star\left(\frac{y-z}{2r}\right)\right\} \\
&= \max_{z\in\mathbb{R}^m} \left\{ g_0(z) - 2rq\left(\frac{y-z}{2r}\right)\right\} \quad (q^\star = q)
\end{aligned}
$$

Proof Cont.

Proof.

$$
\begin{aligned}
g_r(y) &= \inf_{x\in\mathcal{X}} L_r(x,y) \\
&= \inf_{x\in\mathcal{X}} \min_{u\in\mathbb{R}^m} \left(F_r(x,u) + \langle u, y\rangle\right) \\
&= \inf_{u\in\mathbb{R}^m} \left(p_r(u) + \langle u, y\rangle\right) \\
&= -p_r^\star(-y) \quad \textit{(convex conjugate)} \\
&= -(p_0 + 2rq)^\star(-y) \\
&= -(p_0^\star \,\square\, 2rq^\star)(-y) \quad \textit{(infimal convolution)} \\
&= -\inf_{z\in\mathbb{R}^m} \left\{ p_0^\star(-z) + 2rq^\star\left(\frac{y-z}{2r}\right)\right\} \\
&= \max_{z\in\mathbb{R}^m} \left\{ g_0(z) - 2rq\left(\frac{y-z}{2r}\right)\right\} \quad (q^\star = q) \\
&= \max_{z\in\mathbb{R}^m} \left\{ g_0(z) - \frac{1}{4r}\|y-z\|^2\right\}
\end{aligned}
$$

Proof Cont.

Proof.

Therefore,

$$g_r(y) = \max_{z \in \mathbb{R}^m} \left\{ g_0(z) - \frac{1}{4r}\|z - y\|^2 \right\}$$

Note that the expression $g_0(z) - \frac{1}{4r}\|z - y\|^2$ is strongly concave, so the maximal z is unique, and moreover it depends continuously on y.

Proof Cont.

Proof.

Therefore,

$$g_r(y) = \max_{z \in \mathbb{R}^m} \left\{ g_0(z) - \frac{1}{4r} \|z - y\|^2 \right\}$$

Note that the expression $g_0(z) - \frac{1}{4r}\|z - y\|^2$ is strongly concave, so the maximal z is unique, and moreover it depends continuously on y. Since g_0 is not $-\infty$ everywhere, $g_r(y)$ is always finite, and therefore the subgradient is always nonempty. □

Proof Cont.

Proof.

Suppose $w \in \partial g_r(y)$ and z is the maximal z in the defintion of $g_r(y)$. Then,

$$g_0(z) - \frac{1}{4r}\|z - y'\|^2 \leq g_0(z) - \frac{1}{4r}\|z - y\|^2 + \langle w, y' - y \rangle$$

Proof Cont.

Proof.

Suppose $w \in \partial g_r(y)$ and z is the maximal z in the defintion of $g_r(y)$. Then,

$$g_0(z) - \frac{1}{4r}\|z - y'\|^2 \leq g_0(z) - \frac{1}{4r}\|z - y\|^2 + \langle w, y' - y\rangle$$

Equivalently,

$$\frac{1}{4r}\|z - y\|^2 + \langle w, y\rangle \leq \frac{1}{4r}\|z - y'\|^2 + \langle w, y'\rangle$$

This must be true for all $y' \in \mathbb{R}^m$, so it follows that $w = \frac{z-y}{2r}$.

Proof Cont.

Proof.

Suppose $w \in \partial g_r(y)$ and z is the maximal z in the defintion of $g_r(y)$. Then,

$$g_0(z) - \frac{1}{4r}\|z - y'\|^2 \leq g_0(z) - \frac{1}{4r}\|z - y\|^2 + \langle w, y' - y \rangle$$

Equivalently,

$$\frac{1}{4r}\|z - y\|^2 + \langle w, y \rangle \leq \frac{1}{4r}\|z - y'\|^2 + \langle w, y' \rangle$$

This must be true for all $y' \in \mathbb{R}^m$, so it follows that $w = \frac{z-y}{2r}$. Therefore, the subgradient consists of a single element, so $\nabla g_r(y) = \frac{z-y}{2r}$, and therefore $g_r(y)$ is continuously differentiable. □

Proof Cont.

Proof.

Lastly, suppose there exists x such that $L_r(x, y) = g(y)$. Since $L_r(x, y)$ is concave and differentiable in y, we have that

□

Proof Cont.

Proof.

Lastly, suppose there exists x such that $L_r(x, y) = g(y)$. Since $L_r(x, y)$ is concave and differentiable in y, we have that

$$g_r(y') \leq L_r(x, y') \leq L_r(x, y) + \langle \nabla_y L_r(x, y), y' - y \rangle$$

□

Proof Cont.

Proof.

Lastly, suppose there exists x such that $L_r(x, y) = g(y)$. Since $L_r(x, y)$ is concave and differentiable in y, we have that

$$\begin{aligned} g_r(y') \leq L_r(x, y') &\leq L_r(x, y) + \langle \nabla_y L_r(x, y), y' - y \rangle \\ &\leq g(y) + \langle \nabla_y L_r(x, y), y' - y \rangle \end{aligned}$$

□

Proof Cont.

Proof.

Lastly, suppose there exists x such that $L_r(x, y) = g(y)$. Since $L_r(x, y)$ is concave and differentiable in y, we have that

$$\begin{aligned} g_r(y') \le L_r(x, y') &\le L_r(x, y) + \langle \nabla_y L_r(x, y), y' - y \rangle \\ &\le g(y) + \langle \nabla_y L_r(x, y), y' - y \rangle \end{aligned}$$

Therefore, $\nabla_y L_r(x, y) \in \partial g_r(y) = \{\nabla g_r(y)\}$. □

Corollary

Corollary

If g_r is not $-\infty$ everywhere, then for all $y, y' \in \mathbb{R}^m$,

$$g_r(y)+\langle y-y', \nabla g_r(y)\rangle \geq g(y') \geq g_r(y)+\langle y-y', \nabla g_r(y)\rangle - \frac{1}{4r}\|y-y'\|^2$$

Proof.

The first inequality follows from the fact that g_r is concave. For the second inequality, note that for all y there exists $z \in \mathbb{R}^m$ and a quadratic $h(y)$ such that $h(y') = g_0(z) - \frac{1}{4r}\|y' - z\|^2$ with $h(y) = g_r(y)$, and $h(y') \leq g_r(y')$ for all $y' \in \mathbb{R}^m$. This implies that $\nabla h(y) = \nabla g_r(y)$.

Corollary

Corollary

If g_r is not $-\infty$ everywhere, then for all $y, y' \in \mathbb{R}^m$,

$$g_r(y)+\langle y-y', \nabla g_r(y)\rangle \geq g(y') \geq g_r(y)+\langle y-y', \nabla g_r(y)\rangle - \frac{1}{4r}\|y-y'\|^2$$

Proof.

The first inequality follows from the fact that g_r is concave. For the second inequality, note that for all y there exists $z \in \mathbb{R}^m$ and a quadratic $h(y)$ such that $h(y') = g_0(z) - \frac{1}{4r}\|y' - z\|^2$ with $h(y) = g_r(y)$, and $h(y') \leq g_r(y')$ for all $y' \in \mathbb{R}^m$. This implies that $\nabla h(y) = \nabla g_r(y)$. Since h is quadratic, we know that

$$h(y') = h(y) + \langle y - y', \nabla h(y)\rangle - \frac{1}{4r}\|y' - y\|^2$$

Corollary

Corollary

If g_r is not $-\infty$ everywhere, then for all $y, y' \in \mathbb{R}^m$,

$$g_r(y)+\langle y-y', \nabla g_r(y)\rangle \geq g(y') \geq g_r(y)+\langle y-y', \nabla g_r(y)\rangle - \frac{1}{4r}\|y-y'\|^2$$

Proof.

The first inequality follows from the fact that g_r is concave. For the second inequality, note that for all y there exists $z \in \mathbb{R}^m$ and a quadratic $h(y)$ such that $h(y') = g_0(z) - \frac{1}{4r}\|y' - z\|^2$ with $h(y) = g_r(y)$, and $h(y') \leq g_r(y')$ for all $y' \in \mathbb{R}^m$. This implies that $\nabla h(y) = \nabla g_r(y)$. Since h is quadratic, we know that

$$\begin{aligned} h(y') &= h(y) + \langle y - y', \nabla h(y)\rangle - \frac{1}{4r}\|y' - y\|^2 \\ &= g_r(y) + \langle y - y', \nabla g_r(y)\rangle - \frac{1}{4r}\|y' - y\|^2 \end{aligned}$$

Consequences

- We have shown that $g_r(y)$ is (strongly) concave and since $\lim_{r \downarrow 0} L_r(x, y) = L(x, y)$, it follows that it shares the same maximizers as $g_0(y)$

Consequences

- We have shown that $g_r(y)$ is (strongly) concave and since $\lim_{r \downarrow 0} L_r(x, y) = L(x, y)$, it follows that it shares the same maximizers as $g_0(y)$
- We have also computed the gradient of the dual function and shown that it agrees with the gradient of $L_r(x, y)$ in y at the minimizer in x

Consequences

- We have shown that $g_r(y)$ is (strongly) concave and since $\lim_{r \downarrow 0} L_r(x, y) = L(x, y)$, it follows that it shares the same maximizers as $g_0(y)$
- We have also computed the gradient of the dual function and shown that it agrees with the gradient of $L_r(x, y)$ in y at the minimizer in x
- This justifies the construction of the ALM

Consequences

- We have shown that $g_r(y)$ is (strongly) concave and since $\lim_{r\downarrow 0} L_r(x, y) = L(x, y)$, it follows that it shares the same maximizers as $g_0(y)$
- We have also computed the gradient of the dual function and shown that it agrees with the gradient of $L_r(x, y)$ in y at the minimizer in x
- This justifies the construction of the ALM
- After the primal update, the gradients $\nabla g_r(y)$ and $\nabla_y L_r(x, y)$ are close since they are both $\mathcal{C}^1$ and at the minimizer they are equal

Consequences

- We have shown that $g_r(y)$ is (strongly) concave and since $\lim_{r \downarrow 0} L_r(x, y) = L(x, y)$, it follows that it shares the same maximizers as $g_0(y)$
- We have also computed the gradient of the dual function and shown that it agrees with the gradient of $L_r(x, y)$ in y at the minimizer in x
- This justifies the construction of the ALM
- After the primal update, the gradients $\nabla g_r(y)$ and $\nabla_y L_r(x, y)$ are close since they are both $\mathcal{C}^1$ and at the minimizer they are equal
- The perturbed dual problem $(\mathrm{D_r})$ shares the same minimizers as the original dual problem (D), so we do not lose any true solutions or add extraneous solutions from the dual update

Consequences Cont.

- In fact, any KKT pair for (P) and (D) is also a KKT pair for $L_r(x, y)$

Consequences Cont.

- In fact, any KKT pair for (P) and (D) is also a KKT pair for $L_r(x, y)$
- This follows from the above theorem and a very similar proof as before

Consequences Cont.

- In fact, any KKT pair for (P) and (D) is also a KKT pair for $L_r(x, y)$
- This follows from the above theorem and a very similar proof as before
- In fact, we have the following theorem

Consequences Cont.

- In fact, any KKT pair for (P) and (D) is also a KKT pair for $L_r(x, y)$
- This follows from the above theorem and a very similar proof as before
- In fact, we have the following theorem

Theorem

Assume that strong duality holds for (P) *and* (D). *Let* $r > 0$ *and* $y^\star$ *be a dual optimal solution to* (D) *(or equivalently to* ($\mathrm{D_r}$)*). Then,* $x^\star$ *is an optimal solution to* (P) *if and only if* $x^\star = \min_{x \in \mathcal{X}} L_r(x, y^\star) = g_r(y^\star)$.

Proof of Optimality Conditions

Theorem

Assume that strong duality holds for (P) *and* (D)*. Let* $r > 0$ *and* $y^\star$ *be a dual optimal solution to* (D) *or* (D_r)*. Then,* $x^\star$ *is an optimal solution to* (P) *if and only if* $x^\star = \min_{x \in \mathcal{X}} L_r(x, y^\star)$*.*

Proof of Optimality Conditions

Theorem

Assume that strong duality holds for (P) *and* (D). *Let* $r > 0$ *and* $y^\star$ *be a dual optimal solution to* (D) *or* ($\mathrm{D_r}$). *Then,* $x^\star$ *is an optimal solution to* (P) *if and only if* $x^\star = \min_{x \in \mathcal{X}} L_r(x, y^\star)$.

Proof.

The "only if" part is straightforward. Suppose $x^\star = \min_{x \in \mathcal{X}} L_r(x, y^\star)$. Then $x^\star = g_r(y^\star)$ by definition of $g_r(y)$. Moreover, it follows that $\nabla_y L_r(x^\star, y^\star) = \nabla g_r(y^\star) = 0$, so $y^\star$ maximizes $L_r(x^\star, y)$ in y by strong concavity. Thus, $(x^\star, y^\star)$ is a KKT pair and therefore an optimal solution. □

Proof of Optimality Conditions

Theorem

Assume that strong duality holds for (P) *and* (D)*. Let* $r > 0$ *and* $y^\star$ *be a dual optimal solution to* (D) *or* ($\mathrm{D_r}$)*. Then,* $x^\star$ *is an optimal solution to* (P) *if and only if* $x^\star = \min_{x \in \mathcal{X}} L_r(x, y^\star)$*.*

Proof.

The "only if" part is straightforward. Suppose $x^\star = \min_{x \in \mathcal{X}} L_r(x, y^\star)$. Then $x^\star = g_r(y^\star)$ by definition of $g_r(y)$. Moreover, it follows that $\nabla_y L_r(x^\star, y^\star) = \nabla g_r(y^\star) = 0$, so $y^\star$ maximizes $L_r(x^\star, y)$ in y by strong concavity. Thus, $(x^\star, y^\star)$ is a KKT pair and therefore an optimal solution. □

Note that if $r = 0$ then the set of minimizers of $L_0(\cdot, y^\star)$ may points which are not the optimal solution to (P).

Some Definitions

Definition

A **maximizing sequence** of ($\mathrm{D_r}$) is a sequence $\{y^{(k)}\} \subseteq \mathbb{R}^m$ such that $\lim_{k\to\infty} g_r(y^{(k)}) = \sup g_r$.

Definition

A sequence $\{x^{(k)}\} \subseteq \mathcal{X}$ is **asymptotically feasible** for (P) if $\limsup_{k\to\infty} g_i(x^{(k)}) \leq 0$ for $i = 1, \ldots, m$. We define the **asmyptotic optimal value** as the infimum of $\limsup_{k\to\infty} f(x^{(k)})$ over all asymptotically feasible sequences. A sequence $\{x^{(k)}\}$ which converges to this infimum is called **asymptotically minimizing**.

Some Definitions

Definition

A **maximizing sequence** of ($\mathrm{D_r}$) is a sequence $\{y^{(k)}\} \subseteq \mathbb{R}^m$ such that $\lim_{k\to\infty} g_r(y^{(k)}) = \sup g_r$.

Definition

A sequence $\{x^{(k)}\} \subseteq \mathcal{X}$ is **asymptotically feasible** for (P) if $\limsup_{k\to\infty} g_i(x^{(k)}) \leq 0$ for $i = 1, \ldots, m$. We define the **asmyptotic optimal value** as the infimum of $\limsup_{k\to\infty} f(x^{(k)})$ over all asymptotically feasible sequences. A sequence $\{x^{(k)}\}$ which converges to this infimum is called **asymptotically minimizing**.

Remark

The asymptotic optimal value equals the dual optimal value if $d^\star \neq -\infty$ or if there an asmyptotically feasible sequence exists. Therefore, if strong duality holds and (P) is feasible, then $\{x^{(k)}\}$ is asymptotically feasible if and only if $\lim_{k\to\infty} f(x^{(k)}) = p^\star$.

Asymptotic Feasiblility

We provide an example of a family of sequences where no sequence converges to the infimum of the limits.

Asymptotic Feasiblility

We provide an example of a family of sequences where no sequence converges to the infimum of the limits.

- Consider $\{x^{(k,n)}\} = \frac{1}{k} + \frac{1}{n}$

Asymptotic Feasiblility

We provide an example of a family of sequences where no sequence converges to the infimum of the limits.

- Consider $\{x^{(k,n)}\} = \frac{1}{k} + \frac{1}{n}$
- Define $a_n = \lim_{k\to\infty} \left(\frac{1}{k} + \frac{1}{n}\right)$.

Asymptotic Feasiblility

We provide an example of a family of sequences where no sequence converges to the infimum of the limits.

- Consider $\{x^{(k,n)}\} = \frac{1}{k} + \frac{1}{n}$
- Define $a_n = \lim_{k\to\infty} \left(\frac{1}{k} + \frac{1}{n}\right)$.
- Then, $a_n = \frac{1}{n}$ for all n.

Asymptotic Feasiblility

We provide an example of a family of sequences where no sequence converges to the infimum of the limits.

- Consider $\{x^{(k,n)}\} = \frac{1}{k} + \frac{1}{n}$
- Define $a_n = \lim_{k\to\infty} \left(\frac{1}{k} + \frac{1}{n}\right)$.
- Then, $a_n = \frac{1}{n}$ for all n.
- Therefore, $\inf_n a_n = 0$, but no sequence in k converges to 0.

Asymptotic Feasiblility

We provide an example of a family of sequences where no sequence converges to the infimum of the limits.

- Consider $\{x^{(k,n)}\} = \frac{1}{k} + \frac{1}{n}$
- Define $a_n = \lim_{k\to\infty} \left(\frac{1}{k} + \frac{1}{n}\right)$.
- Then, $a_n = \frac{1}{n}$ for all n.
- Therefore, $\inf_n a_n = 0$, but no sequence in k converges to 0.
- Thus, this family of sequences has no asymptotically minimizing sequence.

Convergence Lemma 1

We begin by proving some lemmas required for convergence

Lemma

$$r\|\nabla g_r(y^{(k)})\|^2 \leq \sup g_r - g_r(y^{(k)})$$

Convergence Lemma 1

We begin by proving some lemmas required for convergence

Lemma

$$r\|\nabla g_r(y^{(k)})\|^2 \leq \sup g_r - g_r(y^{(k)})$$

Proof.

From the previous theorems, we know that

$$\sup g_r \geq \max_{y' \in \mathbb{R}^m} \left\{ g_r(y^{(k)}) + \langle (y' - y^{(k)}), \nabla g_r(y^{(k)}) - \frac{1}{4r}\|y' - y^{(k)}\|^2 \right\}$$

$$\geq g_r(y^{(k)}) + \max_{u \in \mathbb{R}^m} \left\{ \langle u, \nabla g_r(y^{(k)}) - \frac{1}{4r}\|u\|^2 \right\}$$

□

Convergence Lemma 1

We begin by proving some lemmas required for convergence

Lemma

$$r\|\nabla g_r(y^{(k)})\|^2 \leq \sup g_r - g_r(y^{(k)})$$

Proof.

From the previous theorems, we know that

$$\begin{aligned}\sup g_r &\geq \max_{y' \in \mathbb{R}^m} \left\{ g_r(y^{(k)}) + \langle (y' - y^{(k)}), \nabla g_r(y^{(k)}) - \frac{1}{4r}\|y' - y^{(k)}\|^2 \right\} \\ &\geq g_r(y^{(k)}) + \max_{u \in \mathbb{R}^m} \left\{ \langle u, \nabla g_r(y^{(k)}) - \frac{1}{4r}\|u\|^2 \right\} \\ &= g_r(y^{(k)}) + r\|\nabla g_r(y^{(k)})\|^2\end{aligned}$$

□

Convergence Lemma 2

Lemma

Suppose the asymptotic optimal value in (P) *is finite. Let* $\{y^{(k)}\}$ *be a bounded maximizing sequence for* $(\mathrm{D_r})$ *with* $r > 0$. *Suppose for each* $x^{(k)} \in \mathcal{X}$

$$L_r(x^{(k)}, y^{(k)}) - \inf_{x\in\mathcal{X}} L_r(x, y_k) = L_r(x^{(k)}, y^{(k)}) - g_r(y^{(k)}) \leq \alpha^{(k)}$$

where $\alpha^{(k)} \to 0$. *Then*

$$r\|\nabla_y L_r(x^{(k)}, y^{(k)}) - \nabla g_r(y^{(k)})\|^2 \leq \alpha^{(k)}$$

Proof of Convergence Lemma

Proof.

Using the corollary proved earlier, we have that

$$L_r(x^{(k)}, y^{(k)}) + \left\langle w - y^{(k)}, \nabla_y L_r(x^{(k)}, y^{(k)}) \right\rangle \geq L_r(x^{(k)}, w)$$

Proof of Convergence Lemma

Proof.

Using the corollary proved earlier, we have that

$$L_r(x^{(k)}, y^{(k)}) + \left\langle w - y^{(k)}, \nabla_y L_r(x^{(k)}, y^{(k)}) \right\rangle \geq L_r(x^{(k)}, w)$$
$$\geq g_r(w)$$

Proof of Convergence Lemma

Proof.

Using the corollary proved earlier, we have that

$$
\begin{aligned}
L_r(x^{(k)}, y^{(k)}) + \left\langle w - y^{(k)}, \nabla_y L_r(x^{(k)}, y^{(k)}) \right\rangle &\geq L_r(x^{(k)}, w) \\
&\geq g_r(w) \\
&\geq g_r(y^{(k)}) + \langle w - y^{(k)}, \nabla_y g_r(y^{(k)}) \rangle - \frac{1}{4r} \| w - y^{(k)} \|^2
\end{aligned}
$$

Proof of Convergence Lemma

Proof.

Using the corollary proved earlier, we have that

$$
\begin{aligned}
L_r(x^{(k)}, y^{(k)}) + \left\langle w - y^{(k)}, \nabla_y L_r(x^{(k)}, y^{(k)}) \right\rangle &\geq L_r(x^{(k)}, w) \\
&\geq g_r(w) \\
&\geq g_r(y^{(k)}) + \langle w - y^{(k)}, \nabla_y g_r(y^{(k)}) \rangle - \frac{1}{4r}\|w - y^{(k)}\|^2
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\alpha^{(k)} &\geq L_r(x^{(k)}, y^{(k)}) - g_r(y^{(k)}) \\
&\geq \sup_{u \in \mathbb{R}^m} \left\{ \left\langle u, \nabla g_r(y^{(k)}) - \nabla_y L_r(x^{(k)}, y^{(k)}) \right\rangle - \frac{1}{4r}\|u\|^2 \right\}
\end{aligned}
$$

Proof of Convergence Lemma

Proof.

Using the corollary proved earlier, we have that

$$
\begin{aligned}
L_r(x^{(k)}, y^{(k)}) + \left\langle w - y^{(k)}, \nabla_y L_r(x^{(k)}, y^{(k)}) \right\rangle &\geq L_r(x^{(k)}, w) \\
&\geq g_r(w) \\
&\geq g_r(y^{(k)}) + \langle w - y^{(k)}, \nabla_y g_r(y^{(k)}) \rangle - \frac{1}{4r} \|w - y^{(k)}\|^2
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\alpha^{(k)} &\geq L_r(x^{(k)}, y^{(k)}) - g_r(y^{(k)}) \\
&\geq \sup_{u \in \mathbb{R}^m} \left\{ \left\langle u, \nabla g_r(y^{(k)}) - \nabla_y L_r(x^{(k)}, y^{(k)}) \right\rangle - \frac{1}{4r} \|u\|^2 \right\} \\
&= r \|\nabla g_r(y^{(k)}) - \nabla_y L_r(x^{(k)}, y^{(k)})\|^2
\end{aligned}
$$

Convergence Theorem

Theorem

Suppose the asymptotic optimal value in (P) *is finite. Let* $\{y^{(k)}\}$ *be a bounded maximizing sequence for* $(\mathrm{D_r})$ *with* $r > 0$. *Suppose for each* $x^{(k)} \in \mathcal{X}$

$$L_r(x^{(k)}, y^{(k)}) - \inf_{x \in \mathcal{X}} L_r(x, y_k) = L_r(x^{(k)}, y^{(k)}) - g_r(y^{(k)}) \leq \alpha^{(k)}$$

where $\alpha^{(k)} \to 0$. *Then* $\{x^{(k)}\}$ *is an asymptotically maximizing sequence for* (P).

Proof of Convergence

From earlier theorems we know that

$$L_r(x^{(k)}, y) = \min_{u \in \mathbb{R}^m} \left\{ F_0(x^{(k)}, u) + \langle u, y \rangle + r\|u\|^2 \right\}$$

where

$$F_r(x, u) = \begin{cases} f(x) + r\|u\|^2 & \text{if } u_i \geq g_i(x), \quad i = 1, \dots, m \\ +\infty & \text{otherwise} \end{cases}$$

Proof of Convergence

From earlier theorems we know that

$$L_r(x^{(k)}, y) = \min_{u \in \mathbb{R}^m} \left\{ F_0(x^{(k)}, u) + \langle u, y \rangle + r\|u\|^2 \right\}$$

where

$$F_r(x, u) = \begin{cases} f(x) + r\|u\|^2 & \text{if } u_i \geq g_i(x), \quad i = 1, \ldots, m \\ +\infty & \text{otherwise} \end{cases}$$

For fixed $y = y^{(k)}$, the minimum is unique, denoted as $u^{(k)}$.

Proof of Convergence

From earlier theorems we know that

$$L_r(x^{(k)}, y) = \min_{u \in \mathbb{R}^m} \left\{ F_0(x^{(k)}, u) + \langle u, y \rangle + r\|u\|^2 \right\}$$

where

$$F_r(x, u) = \begin{cases} f(x) + r\|u\|^2 & \text{if } u_i \geq g_i(x), \quad i = 1, \ldots, m \\ +\infty & \text{otherwise} \end{cases}$$

For fixed $y = y^{(k)}$, the minimum is unique, denoted as $u^{(k)}$.
Therefore, for all $y \in \mathbb{R}^m$,

$$L_r(x^{(k)}, y) \leq F_0(x^{(k)}, u^{(k)}) + \langle u^{(k)}, y \rangle + r\|u^{(k)}\|^2$$

and equality holds when $y = y^{(k)}$.

Proof of Convergence 2

Therefore, for all $y \in \mathbb{R}^m$,

$$L_r(x^{(k)}, y) \leq F_0(x^{(k)}, u^{(k)}) + \langle u^{(k)}, y \rangle + r\|u^{(k)}\|^2$$

and equality holds when $y = y^{(k)}$. Therefore,

$$u^{(k)} = \nabla_y L_r(x^{(k)}, y^{(k)})$$

This implies that $u^{(k)} \to 0$.

Proof of Convergence 2

Therefore, for all $y \in \mathbb{R}^m$,

$$L_r(x^{(k)}, y) \leq F_0(x^{(k)}, u^{(k)}) + \langle u^{(k)}, y \rangle + r\|u^{(k)}\|^2$$

and equality holds when $y = y^{(k)}$. Therefore,

$$u^{(k)} = \nabla_y L_r(x^{(k)}, y^{(k)})$$

This implies that $u^{(k)} \to 0$. To see this, note that by hypothesis

$$\lim_{k\to\infty} L_r(x^{(k)}, y^{(k)}) = \lim_{k\to\infty} g_r(y^{(k)}) = \sup g_r$$

Proof of Convergence 3

To see this, note that by hypothesis

$$\lim_{k\to\infty} L_r(x^{(k)}, y^{(k)}) = \lim_{k\to\infty} g_r(y^{(k)}) = \sup g_r$$

Since $\{y^{(k)}\}$ is bounded, we have that

$$\begin{aligned}\lim_{k\to\infty} F_0(x^{(k)}, u^{(k)}) &= \lim_{k\to\infty} \left(L_r(x^{(k)}, y^{(k)}) - \langle u^{(k)}, y^{(k)}\rangle - r\|u^{(k)}\|^2\right) \\ &= \sup g_r\end{aligned}$$

Remarks

- We can see that the ALM does indeed converge under some mild assumptions.

Remarks

- We can see that the ALM does indeed converge under some mild assumptions.
- However, we did not discuss actual implementations of this method, especially how to solve the subproblem for the primal update

Remarks

- We can see that the ALM does indeed converge under some mild assumptions.
- However, we did not discuss actual implementations of this method, especially how to solve the subproblem for the primal update
- Also, we did not discuss the rate of convergence of the ALM, although this can be found in Rockafellar [2] or Ruszczyński [4, Theorem 6.16]

Conclusion

- The augmented Lagrangian method (ALM) provides a powerful way to deal with constrained optimization problems – even in the nonsmooth case (but we did not cover this)

Conclusion

- The augmented Lagrangian method (ALM) provides a powerful way to deal with constrained optimization problems – even in the nonsmooth case (but we did not cover this)
- We can see the inspiration comes from analyzing the primal and dual problems and attempting to find a pair which satisfies the KKT conditions.

Conclusion

- The augmented Lagrangian method (ALM) provides a powerful way to deal with constrained optimization problems – even in the nonsmooth case (but we did not cover this)
- We can see the inspiration comes from analyzing the primal and dual problems and attempting to find a pair which satisfies the KKT conditions.
- The ALM is of particular interest in conic optimization, especially for semidefinite programs since it may be an alternative to interior-point methods

Conclusion

- The augmented Lagrangian method (ALM) provides a powerful way to deal with constrained optimization problems – even in the nonsmooth case (but we did not cover this)
- We can see the inspiration comes from analyzing the primal and dual problems and attempting to find a pair which satisfies the KKT conditions.
- The ALM is of particular interest in conic optimization, especially for semidefinite programs since it may be an alternative to interior-point methods
- Also, the ALM is not the only primal-dual method, a notable one is the alternating direction method of multipliers (ADMM)

References I

[1] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. doi: 10.1017/CBO9780511804441.

[2] R. T. Rockafellar. Augmented lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of Operations Research*, 1(2):97–116, 1976. ISSN 0364765X, 15265471. URL http://www.jstor.org/stable/3689277.

[3] R. Tyrrell Rockafellar. A dual approach to solving nonlinear programming problems by unconstrained optimization. *Math. Program.*, 5(1):354–373, dec 1973. ISSN 0025-5610. doi: 10.1007/BF01580138. URL https://doi.org/10.1007/BF01580138.

[4] Andrzej Ruszczyński. *Nonlinear Optimization*. Princeton University Press, 2006. ISBN 9780691119151. URL http://www.jstor.org/stable/j.ctvcm4hcj.