arXiv:2302.02072v1 [math.OC] 4 Feb 2023

无限维空间中的非精确偏转次梯度算法

Regina S. Burachik* Xuemei Liu*

2023年2月7日

摘要

我们提出一种对偶方案，用于求解自反巴拿赫空间中的约束非光滑非凸优化问题。对于一类非常一般的增广拉格朗日函数，我们假设增广函数满足较宽松的强制性条件，并证明强对偶性。我们利用这一一般增广拉格朗日函数，通过偏转次梯度方法求解（希尔伯特空间中的）对偶问题。我们为该方法提供了两种步长选择。对于这两种选择，我们均证明原始序列的每个弱聚点都是原始解。我们还证明了对偶序列强收敛。

关键词：增广拉格朗日函数；巴拿赫空间；非凸优化；非光滑优化；次梯度法；对偶方案；罚函数法。

AMS学科分类。49M29；90C25；90C26；90C46；65K10。

1 引言

（广义）增广拉格朗日对偶理论是求解非凸约束优化问题的有力工具。我们无需直接处理约束（原）问题，而是可以通过求解对偶问题来恢复原始解。当对偶问题比原问题更容易求解时，这种方法尤其有用。增广拉格朗日对偶框架便是如此。对偶问题由拉格朗日函数导出，该函数同时包含目标函数和约束条件的信息。强对偶性（即原问题与对偶问题具有相同的最优值）是采用对偶框架的基本要求。然而，对于非凸问题，使用经典拉格朗日函数时，原问题与对偶问题的最优值之间可能存在正的间隙。另一方面，增广拉格朗日对偶 [31,33], 即使在非凸情形下也具有零对偶间隙，并且能够让我们恢复原（非凸）问题的解。接下来，我们将介绍增广拉格朗日函数的起源及其最新进展。

第[31, 章 11] 中引入的线性增广拉格朗日函数，是经典拉格朗日函数与一个增广项之和；换言之，它由目标函数、一个线性项和一个增广项构成。 [31, 例 11.58],中引入的尖锐拉格朗日函数，是一种线性增广拉格朗日函数，它在经典线性项上加上任意范数函数。拉格朗日对偶理论是一个活跃的研究领域，参见例如 [3,4,12,14,19,24,25,28,34–39]。特别是， [37] 和 [38] 针对无限维情形。Dolgopolik 在 [21] 和 [22] 中对不同类型的拉格朗日函数及其在求解各种

*数学系，UniSA STEM，南澳大利亚大学，莫森湖，南澳大利亚州 5095，澳大利亚。电子邮箱：regina.burachik@unisa.edu.au，xuemei.liu@mymail.unisa.edu.au。

问题中的应用进行了研究。更一般类型的拉格朗日函数已在 [18] 以及最近的[16]中得到研究。

我们的目标是使用一个通用且足够简单的拉格朗日函数，为无限维情形提供一个原始–对偶框架。我们的对偶方案与一个算法框架——偏转次梯度方法（DSG）——相结合。已有多项研究在类似的原始–对偶框架中使用 DSG 算法。因此，我们从应用范围（有限维或无限维）、拉格朗日函数和收敛结果等方面，将这些研究与本文工作进行比较。由于这些研究都考察了某种程度的对偶收敛，我们将着重关注原始收敛结果。

Gasimov [23] 提出了一种偏转次梯度算法，该算法使用尖锐拉格朗日函数，其中增广函数为一个范数，用于求解有限维优化问题。此方法具有一个理想特性：它在每次迭代中都会生成一个对偶值严格改进的对偶序列（对偶严格改进）。该方法采用 Polyak 型步长，这要求已知最优对偶值。然而应当注意，这一值在实践中可能难以获得，尤其是在非凸问题中。Gasimov [23] 中的分析仅证明了对偶值序列收敛到最优值。毋庸置疑，原始收敛可能是任何原始–对偶方案最重要的特性，但遗憾的是，[23]并未对此进行研究。事实上，Burachik 等人在 [5, Example1] 中给出的例子表明，[23]中的原始序列可能不会收敛到原始解。

随后，[23], 的工作沿用相同的原始–对偶框架，[5, 7, 15] 进一步研究了步长及收敛结果。[5], Burachik、Gasimov、Ismayilova 和 Kaya 建立了在采用 Polyak 型步长时辅助原始序列的收敛性。在 [7]Burachik、Iusem 和 Melo 提出了 DSG 算法的非精确版本，并证明了非精确迭代下辅助原始收敛。Burachik、Kaya 和 Mammadov [15]设计了 [5, 23] 中方法的非精确版本，并表明即使子问题求解存在一定程度的非精确性，仍可保持相同的收敛性质。

Burachik、Iusem 和 Melo [7] 采用尖锐拉格朗日函数，并提出了两种不依赖于最优值的步长选择。他们证明，采用这些步长时，即使对偶解集为空，原始收敛仍然成立。因此，我们将在本文中采用这类步长，使我们的算法也具备良好的原始收敛性质。

上述四项工作，即 [5, 7, 15, 23],，针对的是有限维情形。现在我们回顾适用于无限维情形的相关工作，这些工作将构成本论文的主要研究动机。

Burachik、Iusem 和 Melo [9] 采用了一种通用形式的增广拉格朗日函数，其中包含了前述四项工作所用的拉格朗日函数作为特例。他们将 [7] 中的分析推广到了无限维情形。具体而言，原问题定义在自反巴拿赫空间中，而约束函数定义在希尔伯特空间中。他们还采用了与 [7], 中相同的 DSG 算法非精确版本，并通过采用 [7] 中的步长类型，确立了原始收敛和对偶收敛。

Burachik 和 Kaya [13], ，以及后来 Burachik、Freire 和 Kaya [11], ，在有限维情形下将一个缩放对称矩阵 A 纳入拉格朗日函数的线性项中。这种通用形式的增广拉格朗日函数将是本文的研究重点，因为它具有足够的普遍性，例如能够对罚函数情形进行完整的理论分析，也就是说，当 A 取为零矩阵时。

我们关注的拉格朗日函数类型是 [11, 13] 中拉格朗日函数向无限维情形的扩展。它与以下无限维等式约束问题相关：

$$\min_{x\in X} \varphi(x) \quad \text{s.t.} \quad h(x) = 0\,, \tag{1}$$

其中 X 是自反巴拿赫空间，$\varphi\colon X \to \mathbb{R}\cup\{\infty\}$ 是下半连续的；并且 $h\colon X \to \mathbb{R}^m$

是连续的。拉格朗日函数 $l\colon X\times\mathbb{R}^m\times\mathbb{R}_+\to\mathbb{R}$ 定义为

$$l(x,y,c):=\varphi(x)-\langle Ay,h(x)\rangle+c\sigma(h(x)),\tag{2}$$

其中 $x\in X$， $y\in\mathbb{R}^m$， $c\in[0,\infty)$， $A\colon\ \mathbb{R}^m\to\mathbb{R}^m$是连续映射，而 $\sigma\colon\ \mathbb{R}^m\to\mathbb{R}_+$ 满足σ（x）$=0$ 当且仅当 $x=0$（更多详情见定义 3.5）。

尖锐拉格朗日函数是式（2）中定义的一般拉格朗日函数的一个特例，其中 A为恒等映射，惩罚函数 σ 是 $\mathbb{R}^m$上的任意范数。此外，当 A 为零[13, 20]时，线性项化为经典惩罚函数。文献[13] 中的数值实验表明，对于不同类别的问题，选择合适的 A 或 σ 可以提升计算性能。

在 [13], 中，采用了 DSG 算法的精确版本，并针对 Polyak 型步长得到了辅助变量和原始问题的收敛结果。我们的分析受到[15] 的启发，用于设计非精确 DSG 算法并建立收敛结果。我们的结果从以下几个方面扩展了 [15] 中的结果。

(i) 我们的拉格朗日函数（2）包含一个一般映射 A，恒等映射和零映射是其特例。关于 A 单位矩阵的情形已在 [15], 中研究，而 $A=0$ 的情形则是新的。这为分析和实现罚函数方法开辟了道路。

(ii) 我们证明了这种非常一般的拉格朗日函数具有强对偶性。特别是，我们考虑的函数σ 可能不是强制的（参见定义 3.4(a′) 和定理 3.1）。

关于原始-对偶框架理论性质的研究，我们指出，[17] 中给出的强对偶性证明也适用于我们的情形。不过，我们在此给出自己的证明，因为我们的拉格朗日函数类型使我们能够得到一个所需假设更弱的结果（更多细节请参见引理 3.1之前的段落）。

在第 3 节中，我们证明我们的无限维框架（i）不存在对偶间隙；（ii）具有一个凸对偶问题，因此可以使用凸分析中的技术求解该问题[1,29,30]。特别是，我们在此引入的算法可视为一种用于最大化对偶函数的 ε-次梯度算法（参见备注 4.3）。

本文结构安排如下。在第 2 节中，我们介绍泛函分析方面的预备知识，以帮助构建原始–对偶框架并建立收敛结果。在第 3 节中，我们给出原始–对偶框架及其重要假设。具体而言，我们将在本节回应上述两个问题，展示该框架的性质并给出相关证明。在第 4 节中，我们陈述偏转次梯度算法。我们提供两种步长选择，并分别建立相应的收敛结果。最后，我们在第 6 节给出结论。

2 预备知识

本章介绍一些供后续使用的泛函分析工具。这些结果大多可以在泛函分析教科书中找到，例如 Brezis [2] 和 Kreyszig [26]的著作。我们以 Brezis 的《泛函分析》[2] 作为主要参考，并为那些未收录于 [2] 其中或难以在其他地方查找到的结果给出自己的证明。这里列出的结果将用于证明原始–对偶框架的性质，以及建立 DSG 算法的收敛结果。

设 X 为自反巴拿赫空间， X^* 为其拓扑对偶（即从 X 到 $\mathbb{R}$的所有连续线性泛函的集合），H 为希尔伯特空间。

我们用 $\langle\cdot,\cdot\rangle$ 表示 $X\times X^*$ 中的对偶积和 H 中的标量积。我们用 $\|\cdot\|$ 表示范数，其中同一记号既用于 X 中的范数，也用于 H 中的范数。我们用记号 $\mathbb{R}_{++}$ 表示正实数， $\mathbb{R}_{+\infty}:=\mathbb{R}\cup\{+\infty\}$（有时简写为 $\mathbb{R}_\infty$，

以及 $\overline{\mathbb{R}} := \mathbb{R}_{+\infty} \cup \{-\infty\}$。给定函数 $g: X \to \overline{\mathbb{R}}$，定义 g 的有效定义域为 $\operatorname{dom} g := \{x \in X: g(x) < +\infty\}$。若 $g(x) > -\infty$ 且 $\operatorname{dom} g \neq \emptyset$，则称 g 为真。回顾一下，集合 $\operatorname{epi} g := \{(x,t) \in X \times \mathbb{R}: g(x) \leq t\}$ 是 g 的上图，而集合 $lev_g(\alpha) := \{x \in X: g(x) \leq \alpha\}$ 是 g 的 α-水平集。设 Y 为巴拿赫空间，并考虑映射 $F: X \to Y$，则 F 的图为集合 $G(F) := \{(x,v) \in X \times Y: v = F(x)\}$。给定 $C \subset X$， C 的示性函数定义为 $\delta_C(v) := 0$，若 $v \in C$；否则为 $+\infty$。若 $C = \{z\}$ 是单点集，则记作 $\delta_{\{z\}} =: \delta_z$。

2.1 泛函分析工具

X（或 H）中由范数诱导的拓扑称为强拓扑。X 中的弱拓扑（H中的弱拓扑）是使 X^* 的所有元素（$H^* = H$的所有元素）连续的最粗拓扑。

定义 2.1（与弱拓扑相关的定义）
设 X 是一个巴拿赫空间， H 是一个希尔伯特空间。

(a) 设 $K \subset X$。如果 K 关于 X 中的弱拓扑是闭的，我们就称 K 在 X 中弱闭。(b) 若对于每个弱开集 $U \subset H$， $h^{-1}(U) \subset X$ 在 X 中都是弱开集，则称函数 $h: X \to H$ 弱-弱连续。(c) 若对于每个强开集 $U \subset H$， $h^{-1}(U) \subset X$ 在 X 中都是弱开集，则称函数 $h: X \to H$ 弱-强连续。强-强连续性和强-弱连续性类似定义。(d) 若函数 $\varphi: X \to \mathbb{R}_{+\infty}$ 关于 X 中的弱拓扑是下半连续（即 w-lsc），则称其弱下半连续（w-lsc）。也就是说， $epi\, \varphi$ 是弱闭的。

接下来我们回顾泛函分析中的一些众所周知的事实。

事实 2.1 设 X 为巴拿赫空间， H 为希尔伯特空间。假设 $K \subset X$ 非空。则有以下结论：如果 $K \subset X$ 弱紧，则它是弱闭的。

在下面的大部分内容中，当某个拓扑性质单独出现时，指的是该性质相对于强拓扑（即范数拓扑）成立。例如，如果我们写“A 是闭的”，我们的意思是“A 是 STRONGLY闭的”。如果某个性质相对于弱拓扑成立，我们会明确提及“弱”或“弱收敛”（例如，弱闭、弱紧等）。

众所周知，在任何度量空间中，紧性等价于序列紧性。为明确巴拿赫空间 X 中弱拓扑情形的性质，我们回顾以下定义。

定义 2.2（弱紧性；序列紧性；强制的）
设 X 为一个巴拿赫空间， $A \subset X$ 和 $\varphi: X \to \mathbb{R}_{+\infty}$。

(a) 当其弱闭包（记为 $\overline{A}^w$）在弱拓扑下为紧集时，集合 A 是弱紧的。(b) 若集合 $A \subset X$ 中的每个序列 $\{x_n\} \subset A$ 都有一个子序列强收敛（分别为弱收敛）到 A 中的某个极限，则称该集合为序列紧（分别为弱序列紧）。

(c) 当 $\lim_{\|x\|\to\infty}\varphi(x) = +\infty$ 时，函数 $\varphi: X \to \mathbb{R}_{+\infty}$ 是强制的。

赋范空间中紧性与序列紧性之间的等价性，使得我们在 X中处理紧集时可以使用序列。为了通过序列来处理 X 中的弱紧集，我们回顾以下经典且著名的结果，该结果出自 [2, 问题 10(3)，第 448] 页。

定理 2.1（埃伯莱因–什穆利安定理）设 E 是一个巴拿赫空间，且 $A \subset E$。令 $B := \overline{A}^w$（即，B 是 A 的弱闭包）。以下陈述等价。

(i) B 是弱紧的。

(ii) B 是弱序列紧的。

接下来，我们引用一些将强拓扑和弱拓扑下的有界性、闭性与紧性联系起来的结果。下一个结果是布尔巴基–阿劳格鲁定理的推论，即 [2, 推论 3.22]。这一结果是凸集的一个分离定理与布尔巴基–阿劳格鲁定理共同推出的。

定理 2.2 设 E 是自反巴拿赫空间。设 $K \subset E$ 是 E 的有界闭凸子集。那么 K 是弱紧的。

推论 2.1 如果 X 是巴拿赫空间，那么每个弱紧集都是闭且有界的。

We 接下来引用一个关于序列紧性的结果，该结果适用于自反巴拿赫空间 .

定理 2.3 设 X 是自反巴拿赫空间，并设 $\{x_n\}$ 是 X 中的有界序列。那么存在一个子序列 $\{x_{n_k}\} \subset \{x_n\}$，它在弱拓扑下收敛。

在我们的分析中，我们将考察弱下半连续函数的水平集。为便于后续使用，我们在下文证明一个可直接由上述引用结果推出的性质。这个性质广为人知，但很难找到与下文表述完全一致的出处。因此，为方便读者，我们在此给出证明。

推论 2.2 设 X 是自反巴拿赫空间，并假设 $\varphi: X \to \mathbb{R}_{+\infty}$ 是弱下半连续的。当且仅当 φ 是强制的，其所有水平集都是弱紧的。在这种情况下，所有水平集都是闭且有界的。

证明。首先假设 φ 是强制的，并固定 $\alpha \in \mathbb{R}$。我们需要证明 $lev_\varphi(\alpha)$是弱紧的。根据埃伯莱因–舒梅连定理（定理2.1），只需证明 $lev_\varphi(\alpha)$是弱序列紧的。这意味着，$lev_\varphi(\alpha)$中的每个序列都包含一个弱收敛到某个极限的子序列，且该极限属于 $lev_\varphi(\alpha)$。事实上，取一个序列 $\{x_k\} \subset lev_\varphi(\alpha)$。由于 φ 是强制的，该序列有界；由定理2.3 可知，存在一个子序列$\{x_{n_k}\}$ 弱收敛到某个 $x \in X$。由于 φ 是弱下半连续的，因此可以写成

$$\varphi(x) \le \liminf_{k\to\infty} \varphi(x_{n_k}) \le \alpha,$$

其中，最后一个不等式成立是因为$x_{n_k} \in lev_\varphi(\alpha)$对所有 k都成立。这意味着 $x \in lev_\varphi(\alpha)$。因此，水平集 $lev_\varphi(\alpha)$是弱序列紧的。根据埃伯莱因–舒梅连定理，它们是弱紧的。反过来，假设对每个 $lev_\varphi(\alpha)$，$\alpha \in \mathbb{R}$都是弱紧的。要证明 φ 是强制的，只需证明 $lev_\varphi(\alpha)$有界。这直接由推论2.1得出，该推论指出，任何弱紧集必定有界。因此，每个水平集$lev_\varphi(\alpha)$都有界，且 φ 是强制的。推论中的最后一个结论直接由推论2.1得出。□

下一个结果对于证明我们将在第 4 节提出的算法的良定义性至关重要。下面引用的结果是 [6, 命题 3.1。15]。

定理 2.4 设 E 为任意拓扑空间，且设 $\varphi: E \to \mathbb{R}_{+\infty}$ 为一个关于 E的拓扑下半连续（lsc）的正常函数。若 $A \subset E$是紧的，且满足 dom $\varphi \cap A \neq \emptyset$，则 φ在 A上有下界，并且能在 A上取得其有限的最小值。

3 原始问题与对偶问题

3.1 理论框架

我们在此提出的原始–对偶框架将 [11, 13] 中研究的框架扩展到了无限维情形。我们对偶框架的一个特例是有限维情形下研究的尖锐拉格朗日函数，参见 [5, 7, 15, 23]。尖锐拉格朗日函数的增广项可以是任意范数，这也是“尖锐”这一术语的由来。在无限维情形下，Burachik、Iusem 和 Melo [9]提出了一种相关的拉格朗日框架。 [9] 中提出的拉格朗日函数使用了惩罚函数$\sigma(\cdot)$，其性质与我们在此研究的相同，但它与我们所用拉格朗日函数的区别在于线性项。具体而言，我们的拉格朗日函数在线性项中包含一个映射 A ，从而为考虑 $A = 0$这一特殊情形下的罚函数方法开辟了途径。 [9]中的框架讨论的是 A 等于恒等映射的情形。尽管我们的一些证明与 [9], 中的证明相似，但由于存在一般映射 A，仍需格外谨慎。由于恒等映射满足我们对映射 A所作的所有假设，因此 [9] 中的所有结果均可由我们的分析推出。为得到本章的主要结果，即强对偶性质，我们采用了一种不同且更为复杂的证明方法。

设 X 是一个自反巴拿赫空间，且 $\varphi: X \to \mathbb{R}_{\infty}$ 是一个正常函数。我们考虑原始优化问题

$$(P) \qquad \min \varphi(x) \text{ s.t. } x \text{ in } X.$$

按照 [17, 第 2 节，2], 我们通过一个函数将问题 (P) 嵌入一族参数化问题中，该函数在参数为零时与目标函数一致。我们使用的工具定义如下。

¯定义 3.1 (P) 的对偶化参数化是一个函数 $f: X \times H \to \mathbb{R}$ ，满足$f(x, 0) = \varphi(x)$，对所有 $x \in X$ 均成立。由该对偶化参数化¯诱导的扰动函数定义为 $\beta: H \to \mathbb{R}$ ，使得

$$\beta(z) := \inf_{x \in X} f(x, z). \tag{3}$$

下一个定义 [17, 定义 5。1], 将成为对偶化参数化的一项基本假设。该定义使用弱开集和弱紧集的概念。我们在定义 2.2(a) 中回顾了后一个概念。另请回想，若一个集合的补集是弱闭集，则该集合为弱开集（见定义 2.1(a)）。

¯定义 3.2 若函数 $f: X \times H \to \mathbb{R}$ 满足：对于每个 $\bar{z} \in H$ 和 $\alpha \in \mathbb{R}$ ，都存在 $\bar{z}$ 的弱开邻域 $U \subset H$ 以及一个弱紧集 $B \subset X$，使得

$$lev_{z,f}(\alpha) := \{x \in X : f(x, z) \leq \alpha\} \subset B \ \ for\ all\ z \in U.$$

换言之，存在弱开集 $U \subset H$ 和弱紧集 $B \subset X$ ，使得 $\bar{z} \in U$，并且我们有

$$\bigcup_{z \in U} lev_{z,f}(\alpha) := \{x \in X : f(x, z) \leq \alpha \ \ \forall z \in U\} \subset B.$$

如果对偶参数化是弱水平紧的，那么相应的扰动函数就是序列弱下半连续的。在证明这一事实之前，我们先回顾其定义。

定义 3.3（序列弱下半连续函数）
令 θ: $H \to \mathbb{R}_{+\infty}$。如果满足以下性质，我们就称 θ 为序列弱下半连续

$$If\ u_n \rightharpoonup u,\ then\ \theta(u) \leq \liminf_{n\to\infty} \theta(u_n).$$

备注 3.1（弱下半连续与序列弱下半连续的区别）
在有限维空间中，或者更一般地，在任意度量空间中，半连续性与其序列版本之间没有区别。然而，在无限维希尔伯特空间中，定义 2.1（d）中给出的弱下半连续性比其序列版本更严格。在前一种情况下，定义 3.3 中的下极限不等式必须对任何弱收敛于某个极限的网成立。由于序列是网的一种特殊情形，弱下半连续性蕴含序列弱下半连续性，反之一般不成立。事实上，定义 2.1（d）对应于上图的弱闭性，而定义3.3 仅对应于该集合是序列弱闭的。

我们将使用以下类型的函数来构造拉格朗日函数。

定义 3.4（增广函数）
函数 $\sigma : H \to \mathbb{R}_{+\infty}$ 是增广函数，当且仅当满足以下性质。

(a) 函数 σ 是真、弱下半连续且强制的（参见定义 *2.1*(d) 和 *2.2*(c)）

(a') 函数 σ 是真、弱下半连续的，并满足以下条件：存在 $K_\sigma > 0$ 使得集合

$$lev_\sigma(K_\sigma) = \{z \in H\ :\ \sigma(z) \leq K_\sigma\},$$

有界。我们称这类 σ 为条件强制的。

(b) 有 $\sigma(0) = 0$ 且 $\operatorname{argmin} \sigma(z) = \{0\}.z$

在下文中，我们始终假设拉格朗日函数中使用的函数 σ 满足定义 3.4中的假设，可以是 (a) 或 (a')。请注意，条件 (a)（强制性）严格强于 (a')（条件强制性）。如果我们能够放宽对 σ的要求，只要求 σ满足条件强制性，我们会在证明中明确说明。否则，我们可以简单地说， σ 满足定义 3.4(a)。在证明强对偶性时，我们将采用对 σ 要求较宽松的假设。在此之前，我们先来证明前面提到的扰动函数的序列弱下半连续性。

命题 3.1（ β 的序列弱下半连续性）
设 f: $X \times Z \to \overline{\mathbb{R}}$ 为弱下半连续且弱水平紧的函数。那么，函数β 定义为 (3) 是序列弱下半连续的。

证明。假设 β 不是序列弱下半连续的。这意味着存在一点 u、一个序列 $\{u_n\}_{n\in\mathbb{N}}$ 以及满足以下条件的 $\varepsilon > 0$

(i) $u_n \rightharpoonup u$, (ii) $\liminf_n \beta(u_n) < \beta(u) - \varepsilon$.

根据 f 的弱层紧性，存在 u 的一个弱邻域 W，使得集合

$$\{x \in X\ :\ f(x,z) \leq \beta(u) - \varepsilon\} \subset B \text{ for all } z \in W,$$

其中， B 在 X中是弱紧的。根据(i)，存在 $n_0 \in \mathbb{N}$ 使得对所有 $n \geq n_0$均有 $u_n \in W$。因此，对所有 $n \geq n_0$，我们有

$$\{x \in X\ :\ f(x,u_n) \leq \beta(u) - \varepsilon\} \subset B.$$

˜将 $F(x)$定义为：$=$上确界$_{n\geq n_0} f(x,u_n)$，这意味着

$$L := lev_{\tilde{F}}(\beta(u)-\varepsilon) = \{x \in X \ : \ \tilde{F}(x) \leq \beta(u)-\varepsilon\} \subset B. \tag{4}$$

˜由于 $f(\cdot,v)$ 对所有 $v \in H$都是弱下半连续的，我们推得 F也是弱下半连续的。因此，集合 L 是弱闭的。由于 L 是弱紧集的弱闭子集，因此它是弱紧的。现在我们可以应用埃伯莱因–舒梅连定理（定理 2.1），从而推出 L 是弱序列紧的。根据 (ii) 以及下极限的定义，我们可以写出

$$\beta(u)-\varepsilon > \sup_{k\in\mathbb{N}} \inf_{n\geq k} \beta(u_n) \geq \inf_{n\geq n_0} \beta(u_n).$$

现在取 $n_1 \geq n_0$，使得对所有 $k \geq n_1$ 都有

$$\beta(u)-\varepsilon-1/k > \inf_{n\geq n_0} \beta(u_n) = \inf_{n\geq n_0} \inf_{x\in X} f(x,u_n),$$

˜其中我们在等式中使用了 β的定义。现在定义 $G(x)$：$=$下确界$_{n\geq n_0} f(x,u_n)$，于是上式变为

$$\beta(u)-\varepsilon-1/k > \inf_{n\geq n_0} \beta(u_n) = \inf_{x\in X} \inf_{n\geq n_0} f(x,u_n) = \inf_{x\in X} \tilde{G}(x),$$

这对所有 $k \geq n_1$都成立。现在固定一个指标 $k \geq n_1$。根据下确界的定义，我们可以找到 $x_k \in \tilde{X}$，使得 $G(x_k)$ $=$下确界$_{n\geq n_0} f(x_k,u_n)$ $<\beta(u)$ $-\varepsilon-1/k$。再次对这个固定的$k \geq n_1$使用类似的论证，我们可以找到 $n_k \geq n_0$和 u_{n_k}，使得 $f(x_k,u_{n_k})$ $<\beta(u)$ $-\varepsilon-1/k$。对每个 $k \geq n_1$执行这一过程，并利用对所有 k 都有 $u_{n_k} \in W$ 这一事实，我们推得所得序列 (x_k) 包含于序列紧集 L 中。因此，存在 (x_k) 的一个子序列，其弱收敛于极限 $\hat{x} \in L$。为简便起见，我们仍将该弱收敛序列记为 (x_k)。因此可以假设 $x_k \rightharpoonup \hat{x} \in L$。由于 $(u_{n_k}) \subset (u_n)$，我们有 $u_{n_k} \rightharpoonup u$。根据 f 的弱下半连续性以及 β 的定义，我们得到

$$\beta(u) \leq f(\hat{x},u) \leq \liminf_k f(x_k,u_{n_k}) \leq \liminf_k \beta(u)-\varepsilon-1/k = \beta(u)-\varepsilon,$$

这与假设矛盾。因此，β 必定是序列弱下半连续的。 □

接下来，我们给出映射 A 所需的基本假设（其中一项涉及增广函数 σ，定义见 3.4）。

假设映射 $A : H \to H$ 满足以下性质：

(A_0) $\sigma(z) \geq \|A(z)\|$，对所有 $z \in H$均成立。

(A_1) 对于 每个 $y \in H$，函数 $\langle A(\cdot)\ y\rangle$ 都是弱上半连续的（即，对于每个 $y \in H$， $-\langle A(\cdot)\ y\rangle$ 都是弱下半连续的 ）.

注 3.2（假设 (A_0)–(A_1)）

(i) 根据定义 3.4 (b)，(A_0) 显然蕴含 $A(0) = 0$。这一假设将在命题 3.2(iii) 中用于推导弱对偶性。假设 (A_0) 对于获得对偶函数的单调不减性质也很重要，正如我们将在命题 3.2(ii) 中看到的那样。

(ii) (A_1) 在引理3.1 和定理3.1中，对于我们的原始–对偶框架实现强对偶性发挥作用。

接下来，我们将列出关于原始–对偶框架的更多假设。

(H0) 目标函数 φ: $X \to \mathbb{R}_{+\infty}$ 是真且弱下半连续。

(H1) 函数 φ 具有弱紧水平集。

（H2）对偶化参数化 f 是真函数（即 dom $f \neq \emptyset$ 和 $f(x, z) > -\infty, \quad \forall (x, z) \in X \times H$），弱下半连续且弱水平紧（参见定义3.2）。

接下来，我们通过增广拉格朗日函数定义与问题（P）对偶的问题。

定义 3.5（增广拉格朗日函数及其相关对偶问题）采用问题（P）的记号，设

(a) f 是定义*3.1*中的对偶化参数化，并满足假设（H2）），

（b） $A: H \to H$ 是满足假设（A_0）–（A_1）的函数。

(c) σ 是定义 *3.4* 中的增广函数，但将 (a) 改为 (a')。

问题 (P) 的增广拉格朗日函数定义为

$$\ell(x, y, c) := \inf_{z \in H} \{f(x, z) - \langle A(z), y\rangle + c\sigma(z)\}. \tag{5}$$

对偶函数 $q: H \times \mathbb{R}_+ \to \mathbb{R}_{-\infty}$ 定义为

$$q(y, c) := \inf_{x \in X} \ell(x, y, c). \tag{6}$$

问题 (P) 的对偶问题为

$$(D) \qquad \text{maximize } q(y, c) \text{ s.t. } (y, c) \in H \times \mathbb{R}_+.$$

分别用 $M_P := \inf_{x \in X} {}^{\varphi}(x)$ 和 $M_D := \sup_{(y,c) \in H \times \mathbb{R}_+} {}^{q}(y, c)$ 表示原始问题和对偶问题的最优值。原始与对偶解集分别记为 $S(P)$ 和 $S(D)$。

注 3.3（问题 (P) 的有限原始值）
根据对偶参数化的定义和假设 (H0)，可知 φ 是真函数，因此根据定义 3.1，有如下结论

$$M_P = \beta(0) < +\infty. \tag{7}$$

3.2 原始–对偶框架的性质

下面介绍定义 3.5中给出的对偶函数的一些基本性质 .

命题 3.2（对偶函数的性质）
设 q 为 (5)中定义的对偶函数。以下结论成立。

(i) 对偶函数 q 是凹函数且弱上半连续（w-usc）。

(ii) 若 $c \geq c_1$，则对于所有 $y \in H$， $q(y, c) \geq q(y, c_1)$。特别地，若 (y, c_1) 是一个对偶解，则对于所有 $c \geq c_1$，(y, c) 也都是对偶解。

(iii) 假设 (A_0) 成立。弱对偶性适用于 $(P) - (D)$ 原始–对偶框架，即

$$M_D = \sup_{(y,c) \in H \times \mathbb{R}_+} q(y, c) \leq \inf_{x \in X} \varphi(x) = M_P,$$

其中 φ 满足 (H0)，且 q 与定义 *3.5*中的定义相同。

证明。（i）我们同时证明 q 是弱上半连续的凹函数。由 (6)， q 是一族弱上半连续凹函数的下确界。事实上，定义 $\psi_{xz}\colon H\times\mathbb{R}\to\mathbb{R}$，使得 $\psi_{xz}(y,c):=f(x,z)-\langle A(z)\ y\rangle+c\sigma(z)$。那么，关于变量 (y,c)，ψ_{xz} 是弱连续且凹的（实际上是仿射的）。由 (6) 可知， q 的凹性和弱上半连续性成立。

现在我们继续证明 (ii)。根据定义以及 $\sigma(z)\geq 0$ 这一事实，可直接得到 $q(y,\cdot)$ 具有单调不减性质。现设 $(y,c_1)\in S(D)$。因此， $q(y,c_1)=M_D$。对于每个 $c\geq c_1$，我们利用单调不减性质可写成

$$M_D\geq q(y,c)\geq q(y,c_1)=M_D,$$

因此，对于所有 $c\geq c_1$， $q(y,c)=M_D$。

现在证明 (iii)。利用公式 (5) 和 (6)，我们有

$$\begin{aligned}M_D=\sup_{(y,c)\in H\times\mathbb{R}_+}q(y,c)&=\sup_{(y,c)\in H\times\mathbb{R}_+}\inf_{x\in X}\inf_{z\in H}\{f(x,z)-\langle A(z),y\rangle+c\sigma(z)\}\\&=\sup_{(y,c)\in H\times\mathbb{R}_+}\inf_{z\in H}\{\inf_{x\in X}f(x,z)\}-\langle A(z),y\rangle+c\sigma(z)\\&=\sup_{(y,c)\in H\times\mathbb{R}_+}\inf_{z\in H}\{\beta(z)-\langle A(z),y\rangle+c\sigma(z)\}\\&\leq\sup_{(y,c)\in H\times\mathbb{R}_+}\{\beta(0)-\langle A(0),y\rangle+c\sigma(0)\}=\beta(0)\\&=\inf_{x\in X}\varphi(x)=M_P,\end{aligned}\tag{8}$$

其中，在第四个等式中我们使用了 β 的定义（见定义 3.1）。我们还使用了 $A(0)=0$ （该结论由 (A_0) 得到）以及 $\sigma(0)=0$ 这一事实（见注释3.2(i) 和定义3.4(b)）。□

我们注意到，上述结果仅要求 σ 来验证定义 3.4 中的性质 (b)。

现在我们着手为我们的原始—对偶框架建立零对偶间隙性质。Burachik 和 Rubinov 在[17]中证明了非常一般的原始—对偶框架下的强对偶性。在他们的分析中，他们使用了抽象凸性 [32]工具。尽管我们可以将零对偶间隙性质作为其分析的一个推论导出，但我们更愿意在此直接证明这一事实。这样做的原因是，在我们的框架中，所使用的一个假设 [17] 可以放宽。具体而言，我们可以将 β 的弱下半连续性假设（用于[17]中）替换为序列-弱下半连续性 β。回顾一下，正如命题 β 3.1所确立的那样，后一个性质对我们的函数成立。

引理 3.1（拉格朗日近似的性质）
考虑原始问题 (P) 及其对偶问题 (D)。假设(H0)–(H2)、(A_0) 以及(A_1)成立。假设增广函数 σ 满足定义 3.4(a')(b)。设存在某个$(\bar{y},\bar{c})\in H\times\mathbb{R}_+$使得 $q(\bar{y},\bar{c})>-\infty$。对于 $n\in\mathbb{N}$，定义 $\gamma_n\colon\ H\to\mathbb{R}_{-\infty}$如下

$$\gamma_n(z):=\beta(z)-\langle A(z),\bar{y}\rangle+n\sigma(z).$$

存在一个序列 $(v_n)\subset H$，具有以下性质。

(i) 存在 $n_0\in\mathbb{N}$，使得

$$\gamma_n(v_n)=\inf_{z\in H}\gamma_n(z),\ \forall\, n\geq n_0.\tag{9}$$

(ii) 序列 $(v_n)\subset H$ 满足 (9) 有界且弱收敛到零。

证明。取满足定理假设的 $(\bar{y},\bar{c})\in H\times\mathbb{R}_+$。令 $\bar{n}:=[\bar{c}]+1$，其中$[\cdot]$ 表示实数的整数部分（或向下取整）。于是 $\bar{n}\in\mathbb{N}$，且由于 $q(\bar{y},\cdot)$ 是递增的

我们有 $q(\bar{y},\bar{n}) \geq q(\bar{y},\bar{c}) > -\infty$。对于任意 $n \in \mathbb{N}$，记 $s(n) := \inf_{z\in H} \gamma_n(z)$。由于$q(\bar{y},\bar{n}) > -\infty$，存在 $r_0 \in \mathbb{R}$ 使得

$$r_0 < q(\bar{y},\bar{n}) = \inf_{z\in H} \beta(z) - \langle A(z),\bar{y}\rangle + \bar{n}\sigma(z) = s(\bar{n}),$$

结合 $(s(n))$ 是一个递增序列这一事实，可得 $s(n) > r_0$ 对所有 $n \geq \bar{n}$成立。从现在起，我们考虑 $n \geq \bar{n}$时的序列 $(s(n))$。注意，该序列 $(s(n))$ 单调递增，下有界于 r_0，上有界于 $\beta(0) = M_P$。事实上，单调性由命题 3.2(ii) 得出。上界的结论直接来自 $s(\cdot)$ 的定义。具体而言，对每个 $n \geq \bar{n}$，都有 $s(n) \leq \gamma_n(0) = \beta(0)$，这里用到了假设 (A_0) 以及 $\sigma(0) = 0$ 这一事实。综上，这些性质意味着序列 $(s(n))$ 收敛（递增地）至极限 $\bar{s} \leq \beta(0)$。

（i）的证明。固定 $n \geq \bar{n}$，我们有

$$s(n) + 1/k > s(n) = \inf_{z\in H} \gamma_n(z),$$

对每个 $k > \bar{n}$ 都成立。根据下确界的定义，我们可以找到 $v_n^k \in H$，使得

$$0 \leq \gamma_n(v_n^k) - s(n) < 1/k, \tag{10}$$

其中，第一个不等式用到了 $s(\cdot)$ 的定义。对每个 $k > \bar{n}$，根据 $s(\bar{n})$ 的定义，我们可以写出

$$s(\bar{n}) \leq \gamma_{\bar{n}}(v_n^k) = \gamma_k(v_n^k) + (\bar{n}-k)\sigma(v_n^k), \tag{11}$$

重新整理可得

$$(k-\bar{n})\sigma(v_n^k) \leq \gamma_k(v_n^k) - s(\bar{n}). \tag{12}$$

根据（10）以及已建立的序列（$s(n)$）的性质，我们知道

$$\gamma_n(v_n^k) < s(n) + 1/k \leq \bar{s} + 1/k < \bar{s} + 1, \tag{13}$$

对所有 $k > \bar{n}$。将（13）代入（12），并重新整理所得表达式，可得

$$\sigma(v_n^k) \leq \frac{(\bar{s} - s(\bar{n})) + 1}{(k-\bar{n})}, \tag{14}$$

其中我们还用到了以下事实：$k > \bar{n}$。现在取 $k_0 := \bar{n} + (\bar{s} - s(\bar{n}) + 1)$，其中 $K_\sigma > 0$ 如定义 3.4(a')所述。接下来可以直接验证

$$\sigma(v_n^k) \leq \frac{(\bar{s} - s(\bar{n})) + 1}{(k-\bar{n})} < K_\sigma,$$

对所有 $k \geq k_0$均成立。现在，由定义3.4(a')可知，集合

$$T := \{v_n^k : n \geq \bar{n}, k \geq k_0\},$$

是有界的。特别地，对于固定的 $n \geq \bar{n}$，序列 $(v_n^k)_{k\geq k_0}$ 是有界的，因此它有一个子序列弱收敛到某个 $v_n \in H$。为简化记号，我们仍将这个弱收敛子列记为 $(v_n^k)_{k\geq k_0}$。根据命题 3.1，我们知道 β 具有序列弱下半连续性。由 (A_1) 和定义 3.4(a')，$-\langle A(\cdot)\ \bar{y}\rangle$ 和 σ 均弱下半连续（因此也具有序列弱下半连续性），由此可推出 γ_n 具有序列弱下半连续性。利用 γ_n 的序列弱下半连续性以及 (13) 中的第一个不等式，我们可以写成

$$\gamma_n(v_n) \leq \liminf_{k\to\infty} \gamma_k(v_n^k) \leq \liminf_{k\to\infty} s(n) + 1/k = s(n) \leq \gamma_n(z),$$

对每个 $z \in H$ 以及每个固定的 $n \geq \bar{n}$ 均成立。上述表达式中的最后一个不等式由 $s(\cdot)$ 的定义得出。令 $n_0 := \bar{n}$，并将 $z = v_n$ 代入上述表达式，即可得到陈述 (i)。

证明 (ii)。现在取在 (i) 部分中定义的序列 (v_n)。注意，在 (i) 部分中定义的集合 T 是有界的，因此存在一个闭球 B_0 使得 $T \subset B_0$。由定理 2.2，B_0 是弱紧的（根据事实 2.1，它也是弱闭的）。这意味着 T 的弱闭包必包含于 B_0。也就是说，

$$\overline{T}^w \subset \overline{B_0}^w = B_0,$$

这表明 $\overline{T}^w$ 是有界的。根据构造（见 (i) 的证明），每个 v_n 都是 T 中某个序列的弱极限，因此我们可以推出

$$(v_n) \subset \overline{T}^w \subset B_0,$$

这表明 (v_n) 是有界的。因此，(ii) 中的有界性结论成立。现在我们继续证明序列 (v_n) 弱收敛于零。为证明这一点，我们将证明每个弱收敛子列都必须收敛于零。若此结论成立，则零是唯一的弱聚点，因此整个序列必弱收敛于零。我们刚刚证明了 (v_n) 是有界的，因此根据定理 2.3，它存在弱收敛子列。任取这样一个子列，记为 $(v_{n_j})_{j\in\mathbb{N}}$，它弱收敛于某个 v。由于 $(v_{n_j}) \subset (v_n)$ 且 $n \geq \bar{n}$，我们可以取 $n_j > \bar{n}$。按照 (11)-(14) 中相同的步骤，以 $k = n_j > \bar{n}$ 和 v_{n_j} 代替 v_n^k，可得

$$\sigma(v_{n_j}) \leq \frac{(\bar{s} - s(\bar{n})) + 1}{(n_j - \bar{n})},$$

由 $n_j \to \infty$。利用 σ 的弱下半连续性，我们可以写出

$$0 \leq \sigma(v) \leq \liminf_{j\to\infty} \sigma(v_{n_j}) \leq \liminf_{j\to\infty} \frac{(\bar{s} - s(\bar{n})) + 1}{(n_j - \bar{n})} = 0,$$

因此 $\sigma(v) = 0$，而关于 σ 的假设推出 $v = 0$。这表明 (v_n) 的每个弱聚点都必为零，因此整个序列弱收敛到零，从而完成了 (ii) 的证明。

□

We 现在我们可以证明原始-对偶框架的强对偶性质 .

定理 3.1（(*P*)–(*D*) 框架的强对偶性）
考虑原始问题 (*P*) 及其对偶问题 (*D*)。假设 (H0)–(H2)、(A_0P) 以及假设增广函数 σ 满足定义 3.4(a')(b)。假设存在某个 $(\bar{y},\ \bar{c}) \in H \times \mathbb{R}_+$ 使得 $q(\bar{y},\ \bar{c}) > -\infty$。则零对偶间隙性质成立，即 $M_P = M_D$。

证明。回顾一下，在我们的设定中弱对偶性（即 $M_D \leq M_P$）成立，这一点已在命题 3.2(iii) 中得到证明。因此，我们只需证明 $M_D \geq M_P$。根据引理 3.1(i)，我们可以取一个满足（9）的序列（v_n）。我们的第一步是证明以下不等式。

$$M_D \geq \liminf_n \gamma_n(v_n), \tag{15}$$

其中 γ_n 和 v_n 与引理 3.1中的相同。利用引理 n_0 3.1(i) 中的 M_D以及 M_D的定义，可得

$$\begin{aligned} M_D &= \sup_{(y,c)\in H\times\mathbb{R}_+} \inf_{z\in H} \{\beta(z) - \langle A(z), y\rangle + c\sigma(z)\} \\ &\geq \inf_{z\in H} \{\beta(z) - \langle A(z), \bar{y}\rangle + n\sigma(z)\} \\ &= \{\beta(v_n) - \langle A(v_n), \bar{y}\rangle + n\sigma(v_n)\} = \gamma_n(v_n), \end{aligned} \tag{16}$$

其中我们在不等式中使用了对 (y,c) 的固定选择：$=(\bar{y},n)$，其中 $n>n_0$；在第二个等式中使用了事实 (9)；在最后一个等式中使用了 γ_n 的定义。对 (15) 取 (16) 中的下极限，即得不等式 (15)。利用 (γ_n 的定义以及下极限的性质，可推出

$$\begin{aligned} M_D &\geq \liminf_n \gamma_n(v_n) \\ &= \liminf_n \beta(v_n) + [-\langle A(v_n), \bar{y}\rangle + n\sigma(v_n)] \\ &\geq \liminf_n \beta(v_n) + \liminf_n [-\langle A(v_n), \bar{y}\rangle + n\sigma(v_n)] \\ &\geq \liminf_n \beta(v_n) + \liminf_n [-\langle A(v_n), \bar{y}\rangle] \\ &\geq \beta(0) + 0 = \beta(0) = M_P, \end{aligned} \tag{17}$$

其中，第三个不等式用到了以下事实：$n\sigma(v_n) \geq 0$。在最后一个不等式中，我们用到了引理3.1(ii)，即事实（v_n）弱收敛于零，以及以下事实：$A(0)=0$。更准确地说，利用函数 β 和 $-\langle A(\cdot)\ \bar{y}\rangle$的（序列）弱下半连续性，可得

$$\liminf_n [-\langle A(v_n), \bar{y}\rangle] \geq -\langle A(0), \bar{y}\rangle = 0,$$

and

$$\liminf_n \beta(v_n) \geq \beta(0).$$

这两个事实都用在了式（17）最后一个不等式中。由于我们已经有 $M_D \leq M_P$，因此已证明 $M_P = M_D$。
□

定义 3.6（凹函数的超微分）
设 H 为一个希尔伯特空间，$g: H \to \mathbb{R}_{-\infty}$ 为一个凹函数。取 $r \geq 0$。g 在 $w_0 \in$ 处的 r-超微分，其中 $w_0 \in$ 属于 dom(g): $= \{w \in H: g(w) > -\infty\}$，是由下式定义的集合 $\partial_r g(w_0)$:

$$\partial_r g(w_0) := \{v \in H : g(w) \leq g(w_0) + \langle v, w - w_0\rangle + r, \ \forall v \in H\}.$$

定义 3.7（原始–对偶解与拉格朗日函数的近似）我们称

(i) 若 $\varphi(x_*) \leq M_P + \epsilon$，则称 $x_* \in X$ 为 (P) 的 ϵ-最优原始解

(ii) 若 $q(y_*, c_*) \geq M_D - \epsilon$，则称 $(y_*, c_*) \in H \times \mathbb{R}_+$ 为 ϵ-最优对偶解。

(iii) 对于 $r \geq 0$，定义集合

$$X_r(y,c) := \{(x,z) \in X \times H : f(x,z) - \langle A(z), y\rangle + c\sigma(z) \leq q(y,c) + r\}, \tag{18}$$

它包含增广拉格朗日函数的所有 r-极小化点。

¯(iv) 固定 $(w,c) \in H \times \mathbb{R}_+$，并定义 $\Phi_{(w,c)}: X \times H \to \mathbb{R}$ 如下

$$\Phi_{(w,c)}(x,z) := f(x,z) - \langle A(z), w\rangle + c\sigma(z). \tag{19}$$

注 3.4（拉格朗日函数逼近的对偶集）
根据 q 作为下确界的定义，对于每个 $r>0$ 以及满足 $q(y,c) > -\infty$ 的任意 (y,c)，存在 (x,z) 使得 $f(x,z) - \langle A(z)\ y\rangle + c\sigma(z) < q(y,c) + r$。因此，对于每个 $r>0$ 以及满足 $q(y,c) > -\infty$ 的任意 (y,c)，我们有 $X_r(y,c)$ 非空。

下面的结果推广了 [9, 命题 3.1 的第 (i) 和 (iii) 部分]，其中考虑了 $A=I$，即 H 中的恒等映射。由于证明只需对 [9, 命题 3.1 的第 (i) 和 (iii) 部分]，中的相同步骤稍作修改，因此省略证明。

命题 3.3 若 $(\hat{x}, \hat{z}) \in X_r(\hat{y}, \hat{c})$，则以下结论成立。
i) 对所有 $r \geq 0$，$(-A(\hat{z}) \ \sigma(\hat{z})) \in \partial_r q(\hat{y}, \hat{c})$。
ii) 如果 $M_D \leq M_P$ 且 $\hat{z} = 0$，则 $\hat{x}$ 是一个 r-最优原始解，并且 $(\hat{y}, \hat{c})$ 是一个 r-最优对偶解。特别地，如果 $r \leq \epsilon$，则 $\hat{x}$ 是一个 ϵ-最优原始解，并且 $(\hat{y}, \hat{c})$ 是一个 ϵ-最优对偶解。

从现在起，我们假设定理 **3.1** 的假设均得到验证，因此有 $M_P = M_D$。

下面的结果建立了原始–对偶解集的若干性质，以及函数 $\Phi_{(y,c)}$ 在（19）中定义的水平集的紧性性质。关于 (i)、(ii)、（20）中集合的非空性以及 (iiiA) 的证明技术都是标准的，可在 [9, 引理 3。1]中找到。因此，我们将省略这些证明。然而，(iiiB) 的证明是新的，因为对 σ的强制性假设（该假设用于 [9], ）放宽为定义 3.4 中带有条件 (a') 的较弱版本。因此，我们在此给出该部分的证明。

定理 3.2（拉格朗日函数的紧水平集）
考虑原始问题 (P) 及其对偶问题 (D)。假设 (H0)–(H2)、(A_0) 和 (A_1) 成立。则以下结论成立。

(i) 集合 $S(P) \neq \emptyset$ 以及 $M_P \in \mathbb{R}$。

(ii) 令 $(\hat{y}, \hat{c}) \in H \times \mathbb{R}_+$ 满足 $q(\hat{y}, \hat{c}) > -\infty$，并考虑集合

$$T := \{(w, c) \in H \times \mathbb{R}_+ \ : \ c > \hat{c} + \|w - \hat{y}\|\}.$$

那么，

(iiA) $T \subset \text{dom } q$，即对于每个 $(w, c) \in T$，都有 $q(w, c) > -\infty$。(iiB) 若 $(\hat{y}, \hat{c}) \in S(D)$，则 $T \subset S(D)$。

(iii) 对于每个 $s \geq M_P$ 以及每个 $(w, c) \in H \times \mathbb{R}_+$，水平集

$$lev_{\Phi_{(w,c)}}(s) = \{(x, z) \in X \times H : \Phi_{(w,c)}(x, z) = f(x, z) - \langle A(z), w\rangle + c\sigma(z) \leq s\}, \tag{20}$$

非空。

(iiiA) 令 $(\hat{y}, \ \hat{c})$ 如 (ii) 中所述，并假设 σ 满足定义 *3.4* 中的条件 (a)。则对于每个 (20)，其中的水平集都是弱紧的$(w, \ c) \in T$。在这种情况下，存在$(\tilde{x}, \ \tilde{z})$ 使得

$$q(w, c) = f(\tilde{x}, \tilde{z}) - \langle A(\tilde{z}), w\rangle + c\sigma(\tilde{z}). \tag{21}$$

(iiiB) 令 $(\hat{y}, \ \hat{c})$ 如 (ii) 中所述，并假设 σ 满足定义 *3.4* 中的条件 (a')。定义集合

$$\tilde{T}(s) := \{(w, c) \in H \times \mathbb{R}_+ \ : \ c > \hat{c} + \left(\frac{s - q(\hat{y}, \hat{c})}{K_\sigma}\right) + \|w - \hat{y}\|\} \subset T,$$

其中 $K_\sigma > 0$ 如定义 *3.4*(a') 中所述。那么，(20) 中的水平集是弱紧的˜对每个 (w, c) $\in T(s)$均成立。在这种情况下，存在 $(\tilde{x}, \tilde{z})$ 使得 (21) 成立。

证明。(i)、(ii)、(20) 中集合的非空性以及 (iiiA) 的证明与 [9, 引理 31] 类似。下面证明 (iiiB)。首先注意，$\tilde{T}(s) \subset T$，因为 $\left(\frac{s-q(\hat{y},\hat{c})}{K_\sigma}\right) \geq 0$。事实上，注意 $s \geq M_P = M_D \geq q(\hat{y}, \hat{c})$ 且 $K_\sigma > 0$。接下来要证明，在 (iiiB) 给出的假设下，$lev_{\Phi_{(w,c)}}(s)$ 弱紧。具体而言，我们需要证明水平集

˜在 (20) 中对每个 $(w,c) \in T(s)$ 都是弱紧的。根据定理 2.1，只需证明集合 $lev_{\Phi_{(w,c)}}(s)$ 弱序列紧。也就是说，包含于 $lev_{\Phi_{(w,c)}}(s)$ 中的每个序列都有一个弱收敛子列，且该弱收敛子列的极限属于 $lev_{\Phi_{(w,c)}}(s)$。取一个序列 $\{(x_k,z_k)\} \subset lev_{\Phi_{(w,c)}}(s)$。我们首先证明 $\{z_k\}$ 有一个弱收敛子列。事实上，根据 $lev_{\Phi_{(w,c)}}(s)$ 的定义，我们有

$$\begin{aligned} s &\geq f(x_k,z_k) - \langle A(z_k), w\rangle + c\sigma(z_k) \\ &= f(x_k,z_k) - \langle A(z_k), \hat{y}\rangle + \hat{c}\sigma(z_k) + \langle A(z_k), \hat{y} - w\rangle + (c-\hat{c})\sigma(z_k) \\ &\geq f(x_k,z_k) - \langle A(z_k), \hat{y}\rangle + \hat{c}\sigma(z_k) - \|A(z_k)\|\|\hat{y} - w\| + (c-\hat{c})\sigma(z_k) \\ &\geq q(\hat{y},\hat{c}) + (c - \hat{c} - \|w-\hat{y}\|)\sigma(z_k), \end{aligned}$$

其中，第二个不等式中用到了柯西-施瓦茨不等式，而第三个不等式中用到了 q 和(A_0)的定义。由于 $T(s) \subset T$且$(w,c) \in T(s)$，我们有$(w,c) \in T$，因此$(c-\hat{}c-\|w-\hat{}y\|) > 0$。事实 $q(\hat{}y,\hat{}c) > -\infty$，结合 φ 的真性，可推出 $q(\hat{}y,\hat{}c) \in \mathbb{R}$。综上，我们可以重新整理最后一个表达式，得到

$$\sigma(z_k) \leq \frac{s - q(\hat{y},\hat{c})}{c - \hat{c} - \|w-\hat{y}\|} =: M(c).$$

˜由于 $s \geq M_P = M_D \geq q(\hat{}y,\hat{}c)$，我们有 $M(c) \geq 0$。现在我们将用到事实$(w,c) \in T(s)$。事实上，这一假设意味着

$$c > \frac{s - q(\hat{y},\hat{c})}{K_\sigma} + \hat{c} + \|w-\hat{y}\|.$$

在对 (w,c) 作出此假设后，直接可验证 $M(c) < K_\sigma$。根据定义 3.4(a')，序列 $\{z_k\}$ 有界，因此存在一个弱收敛子列。不失一般性，我们可以假设整个序列 $\{z_k\}$ 弱收敛于某个 $\bar{z}$。现在我们来寻找 $\{x_k\}$ 的一个弱收敛子列。事实上，利用以下事实$\{(x_k,z_k)\} \subset lev_{\Phi_{(w,c)}}(s)$，我们可以写出

$$f(x_k,z_k) \leq s + \langle A(z_k), w\rangle - c\sigma(z_k) \leq s + \|w\|\|A(z_k)\| \leq s + \|w\|\sigma(z_k) \leq s + \|w\|M(c) =: \tilde{\alpha} \quad (22)$$

˜对于某个 $\tilde{\alpha} \in \mathbb{R}$(注意，$(w,c) \in T(s)$ 是固定的)。由 f的弱水平紧性（见定义3.2），存在一个弱紧集 $B \subset X$ 和一个 U 的弱开邻域 $\bar{z}$，使得

$$\bigcup_{z\in U} lev_{z,f}(\tilde{\alpha}) \subset B,$$

其中 $lev_{z,f}(\tilde{\alpha}) := \{x \in X\colon\ f(x,z) \leq \tilde{\alpha}\}$。由于 $\{z_k\}$弱收敛于 $\bar{z}$，且 U是弱开集，因此存在一个 k_0，使得 $z_k \in U$ 对所有 $k > k_0$均成立。利用 (22)，我们推得

$$\{x_k\}_{k>k_0} \subset \bigcup_{k>k_0} \{x \in X\ :\ f(x,z_k) \leq \tilde{\alpha}\} \subset B.$$

因此，$\{x_k\}_{k>k_0} \subset B$ 并且由于 B 是弱紧的，$\{x_k\}_{k>k_0}$存在一个子序列弱收敛于某个 $\bar{x}$。总之，我们已经证明，$\{(x_k,z_k)\}$ 存在一个弱收敛子序列 $\{(x_{k_j},z_{k_j})\}$，其极限为$(\bar{x},\bar{z})$。回顾一下，f和 σ 是弱下半连续的，而函数 $\langle A(\cdot)\ w\rangle$ 是弱上半连续的。因此，$-\langle A(\cdot)\ w\rangle$ 是弱下半连续的。总之，由式（19）给出的函数 $\Phi_{(w,c)}$ 是弱下半连续的。对于弱收敛子序列 $\{(x_{k_j},z_{k_j})\}$，我们可以写成

$$\Phi_{(w,c)}(\bar{x},\bar{z}) \leq \liminf_{j\to\infty} \Phi_{(w,c)}(x_{k_j},z_{k_j}) \leq s,$$

其中最后一个不等式由假设 $\{(x_k,z_k)\} \subset lev_{\Phi_{(w,c)}}(s)$ 得出。因此，我们证明了弱极限 $(\bar{x},\bar{z})$ 属于 $lev_{\Phi_{(w,c)}}(s)$，故后者是弱紧的，如

前所述。现在我们继续证明 (iiiB) 中的最后一个结论，这需要存在如 (21) 所示的 $(\tilde{x},\tilde{z})$。我们首先注意到，根据 (ii) 以及 φ 的真性，对于每个 $(w,c)\in T,\tilde{}$，$q(w,c)\in\mathbb{R}$，因此对于每个 $(w,c)\in T$ (s)，同样成立。为了证明 (21) 中的等式，我们需要说明对应于值 $q(w,c)$ 的下确界确实能取到。我们断言，(21) 中的等式可由以下事实推出：

$$\operatorname*{argmin}_{(x,z)\in X\times H}\Phi_{(w,c)}(x,z)\neq\emptyset. \tag{23}$$

事实上，假设（23）成立，并取（$\tilde{x}$，$\tilde{z}$）$\in\operatorname{argmin}_{(x,z)\in X\times H}\Phi_{(w,c)}$（$x$，$z$）。因此，

$$q(w,c)=\inf_{(x,z)\in X\times H}\Phi_{(w,c)}(x,z)=\Phi_{(w,c)}(\tilde{x},\tilde{z})=f(\tilde{x},\tilde{z})-\langle A(\tilde{z}),w\rangle+c\sigma(\tilde{z}), \tag{24}$$

其中我们用到了 q 和 $\Phi_{(w,c)}$的定义，以及关于 $(\tilde{x},\tilde{z})$ 的假设。因此，只需证明（23），等式（21）便成立。我们知道（20）中的集合非空，并且已经证明它是弱紧的。采用定理 2.4 的记号，令 $E:=X\times H$，并在E 中考虑弱拓扑，令 $\varphi:=\Phi_{(w,c)}$ 且 $K:=lev_{\Phi_{(w,c)}}(s)$。根据水平集的定义，有 $lev_{\Phi_{(w,c)}}(s)\subset\operatorname{dom}\Phi_{(w,c)}$。综上，我们有

$$\operatorname{dom}\Phi_{(w,c)}\cap lev_{\Phi_{(w,c)}}(s)=lev_{\Phi_{(w,c)}}(s)\neq\emptyset,$$

其中，非空性由第 (iii) 部分的第一条陈述得出。由于 $\Phi_{(w,c)}$ 是弱下半连续的，定理 2.4 的所有假设均成立，因此 $\Phi_{(w,c)}$ 在集合 $lev_{\Phi_{(w,c)}}$（s）上有下界，并且在该集合上达到其最小值。这就证明了（23），定理的证明至此完成。□

4 偏转次梯度算法（DSG）

本文将始终使用以下记号。

$$q_k:=q(y_k,c_k),$$
$$\bar{q}:=q(\bar{y},\bar{c}),$$

其中，$(\bar{y},\bar{c})$ 表示一个对偶解，因此对于每个 k，$\bar{q}=q(\bar{y},\bar{c})=M_D=M_P\geq q_k$。

4.1 定义与收敛分析

本节定义偏转次梯度算法（DSG），并确立其收敛性质。我们首先定义偏转次梯度算法（DSG）。

算法 4.1 偏转次梯度算法（DSG）

步骤 0。选择 $(y_0,c_0)\in H\times\mathbb{R}_+$ 使得 $q(y_0,c_0)>-\infty$，以及外生参数$\epsilon>0$ （预设容差）、$\delta<1$、$\{\alpha_k\}\subset(0,\alpha)$，其中 $\alpha>0$，并选择 $\{r_k\}\subset\mathbb{R}_+$使得 $r_k\to 0$。令 $k:=0$。

步骤 1。（子问题和停止准则）

a) 求解 $(x_k,z_k)\in X_{r_k}(y_k,c_k)$,

b) 如果 $z_k=0$ 且 $r_k\leq\epsilon$，则停止,

c) 如果 $z_k=0$ 且 $r_k>\epsilon$，则 $r_k:=\delta r_k$ 并转到 (*a*),

d) 如果 $z_k\neq 0$，则转到步骤 2。

步骤 2.（步长选择与变量更新）

考虑一个步长 $s_k>0$，并定义

$y_{k+1} := y_k - s_k A(z_k), c_{k+1} :$
$= c_k + (\alpha_k + 1) s_k \sigma(z_k), k :$
$= k + 1$，转到步骤 1。

注 4.1 注意，当 $A = 0$时，偏转次梯度算法就成为经典罚函数法。对于 $A := I$ 在 H中的恒等映射，我们得到 [9, 第 3] 节中定义的 IMSg算法。

注 4.2 根据注 3.4，当 $r_k > 0$时，存在如步骤 1(a) 中所示的 $(x_k, z_k) \in X_{r_k}(y_k, c_k)$，这表明该算法的非精确版本总是良定义的。当 $r_k = 0$ 对所有 k都成立时，我们得到精确版本；该版本在第一个满足 k 的 $z_k = 0$ 处停止。精确版本的良定义性将在命题 4.2中证明。在后一个结果中，我们给出精确版本的步骤 1(a) 可执行的条件。我们的分析包括对 σ 的选择，该选择可以如定义 3.4的 (a) 部分或 (a') 部分所述。

注 4.3 利用命题 3.3，我们可以看到，算法 4.1 中的步骤 2 只不过是最大化对偶函数的一个 ε-次梯度步。

注 4.4 若 $z_k \in N(A)$，则 $y_{k+1} = y_k$ 且 $c_{k+1} > c_k$。因此，在这种情况下不会更新 y_k。$z_k \in N(A)$ 的情况不会对收敛造成问题。事实上，我们的结果适用于 $A = 0$，因此 $N(A) = X$。

接下来，我们给出对每个步长 $s_k > 0$ 以及每个 $\alpha_k \in (0, \alpha)$ 都成立的性质。命题 (a) 和 (b) 等价的证明是标准的，只需稍作修改即可沿用 [9, 命题3.1(ii)]的相同步骤。不过，我们在此给出简短的证明，因为证明中涉及的表达式将在后续结果中频繁用到。

命题 4.1（对偶收敛的刻画）设 $\{(x_k, z_k)\}$ 和 $\{(y_k, c_k)\}$ 是偏转次梯度算法生成的序列，并假设 (A_0) 成立。以下陈述等价。

(a) 对偶序列 $\{(y_k, c_k)\}$ 有界。(b) $\sum_k s_k \sigma(z_k) < +\infty$.(c) 对偶序列 $\{(y_k, c_k)\}$ 强收敛到某个极限。(d) 序列 $\{c_k\}$ 是柯西序列。(e) 序列 $\{c_k\}$ 有界。

此外，若 $\{c_k\}$ 有界，则 $\{y_k\}$ 也有界。

证明。利用 (A_0) 和 $\{y_k\}$ 的定义，可得

$$\|y_{k+1} - y_0\| \le \sum_{j=0}^{k} \|y_{j+1} - y_j\| = \sum_{j=0}^{k} s_j \|A(z_j)\| \le \sum_{j=0}^{k} s_j \sigma(z_j). \tag{25}$$

另一方面，根据 $\{c_k\}$ 的定义，我们有

$$c_{k+1} - c_0 = \sum_{j=0}^{k} c_{j+1} - c_j = \sum_{j=0}^{k} (\alpha_j + 1) s_j \sigma(z_j) \le (\alpha + 1) \sum_{j=0}^{k} s_j \sigma(z_j), \tag{26}$$

其中我们用到了以下事实： $\alpha_k < \alpha$ 对每个 k。我们先证明 (a) 与 (b) 等价。若 (b) 成立，则由 (25) 和 (26) 可立即推出 (a)。反之，假设 (a) 成立。利用 (26) 的左端以及 (a)，可知存在 $M > 0$ 使得

$$M \geq c_{k+1} - c_0 = \sum_{j=0}^{k} c_{j+1} - c_j = \sum_{j=0}^{k} (\alpha_j + 1) s_j \sigma(z_j) \geq \sum_{j=0}^{k} s_j \sigma(z_j), \tag{27}$$

这里我们用到了以下事实：对于所有 j， $\alpha_j > 0$ 。由于上述不等式对所有 k都成立，因此必有 $\sum_k^{\infty} s_k \sigma(z_k) \leq M$ ，由此我们得出 (b)。现在来证明 (a) 等价于 (c)。显然，(c) 蕴含 (a)，因此只需证明 (a) 蕴含 (c)。假设 (a) 成立。那么序列 $\{c_k\}$ 有界。由于它严格递增，因此必定收敛。特别地，这意味着序列 $\{c_k\}$是柯西序列。现在我们将证明，关于范数， $\{y_k\}$ 也是柯西序列。事实上，对于每个 k、 $j \in \mathbb{N}$ ，我们可以写出

$$c_{k+j} - c_k = \sum_{l=k}^{k+j-1} c_{l+1} - c_l = \sum_{l=k}^{k+j-1} (\alpha_l + 1) s_l \sigma(z_l) \geq \sum_{l=k}^{k+j-1} s_l \sigma(z_l) \geq \|y_{k+j} - y_k\|, \tag{28}$$

这里我们再次用到了以下事实： $\alpha_j > 0$ 对于所有 j 在第一个不等式中。最后一个不等式的推导与 (25) 类似，只需将 $k + j - 1$ 替换为 k，并将 k 替换为 0。由于 $\{c_k\}$ 是柯西序列，因此对于所有 $j \in \mathbb{N}$ ，我们有

$$0 = \lim_{k \to \infty} c_{k+j} - c_k \geq \lim_{k \to \infty} \|y_{k+j} - y_k\| \geq 0, \tag{29}$$

因此 $\{y_k\}$ 也如所声称的那样是柯西序列。我们的断言成立，并且由于 H 是完备的，序列 $\{y_k\}$强收敛到极限 $\bar{y}$。总之， $\{(y_k, c_k)\}$ 强收敛，因此 (a) 推出 (c)。由于 (c) 推出 (d)，要完成证明，只需证明 (d) 推出 (c)。这与证明 (a) 推出 (c) 的方法类似。事实上，若 $\{c_k\}$ 是柯西序列，则由 (29) 可推得 $\{y_k\}$ 也是柯西序列，因此由 $H \times \mathbb{R}$的完备性可知 (c) 成立。显然，(d) 推出 (e)。若 (e) 成立，则由 (26) 和 (25) 可知 $\{y_k\}$也有界，因此 (a) 成立。证明完毕。□

接下来的结果证明该算法是良定义的，也就是说，步骤 1 中执行的极小化问题存在解。我们证明，当 σ 是强制的，或满足定义3.4(a') 中的条件时，这一事实成立。接下来结果中 (i) 部分的证明沿用 [9, 命题 3.2], 的步骤，因此省略其证明。(ii) 部分采用了对 σ较弱的假设，即条件强制性。

˜命题 4.2（DSG 的良定义性）设 $s \geq M_P$，$(\hat{}y,\ \hat{}c)$，以及集合 T、$T(s)$ 如定理 *3.2* 中所述。

(i) 假设 σ 满足定义 *3.4*(a)。取 $y_0 : = \hat{}y$ 和 $c_0 > \hat{}c$。则由 $\{(y_k,\ c_k)\}$ 生成的精确版 DSG 对偶序列 $(y_0,\ c_0)$ 是良定义的。也就是说，集合

$$X(y_k, c_k) := \{(x, z) \in X \times H : f(x, z) - \langle A(z), y_k \rangle + c_k \sigma(z) = q(y_k, c_k)\},$$

对所有 $k \geq 0$ 都非空。

(ii) 假设 σ 满足定义 *3.4*(a')。取 $y_0 := \hat{}y$ 和 $c_0 > \hat{}c + {}^{s-q}(\hat{y},\ \hat{c})\ K_\sigma$ 。则 (i) 中的相同结论成立。

特别是，只要针对相应类型的 σ，按 (i) 或 (ii) 选择 (y_0, c_0)，精确版 DSG 就是良定义的；也就是说，对于所有 $k \geq 1$，都存在 $(x_k, z_k) \in X \times H$满足 $q(y_k, c_k) = [f(x_k, z_k) - \langle A(z_k)\ y_k \rangle + c_k \sigma(z_k)] \in \mathbb{R}$ 对每个 $k \geq 1$ 均成立。

˜证明。(i) 与 [9, 命题 3.2] 类似。(ii) 根据假设 $(y_0, c_0) \in T(s)$，由定理 3.2(iiiB) 可知，存在 $(x_0, z_0) \in X(y_0, c_0)$。如果 $z_0 = 0$ 算法在 $k = 0$ 处停止，那么 (ii) 中的结论对唯一的迭代点 (y_0, c_0) 成立。假设 $k \geq 1$ （即算法未在$k = 0$处停止，因此 $z_0 \neq 0$）。下面证明，对每个 $k \geq 1$，都有 $(y_k, c_k) \in T(s)$。由 (25)，我们有 $\sum_{j=0}^{k-1} s_j \sigma(z_j) \geq \|y_k - y_0\|$；由 (26)，我们有 $c_k - c_0 = \sum_{j=0}^{k-1}(\alpha_j + 1)s_j\sigma(z_j)$。综上可得

$$c_k = c_0 + \sum_{j=0}^{k-1}(\alpha_j + 1)s_j\sigma(z_j) \geq c_0 + \|y_k - y_0\| + \sum_{j=0}^{k-1}\alpha_j s_j \sigma(z_j)$$

$$> c_0 + \|y_k - y_0\| > (\hat{c} + \frac{s - q(\hat{y}, \hat{c})}{K_\sigma}) + \|y_k - y_0\|,$$

其中，第一个严格不等式源于以下事实：$z_j \neq 0$ （等价地，$\sigma(z_j) \neq 0$）对每个$0 \leq j \leq k-1$ 都成立（否则算法将在某个 $j < k$处停止），而第二个严格˜不等式使用了 c_0的定义。因此，$(y_k, c_k) \in T(s)$ 对每个 $k \geq 1$ 都成立，结论由定理3.2(iiiB)得出。该命题的最后一个断言直接由(i)和(ii)得出。□

以下结果的(a)部分已在 $A = I$ 和 σ 在[9, 中为强制的情形下，于引理3中得到证明。2], 它在不对偏转次梯度算法的参数作任何额外假设的情况下，确立了序列 $\{z_k\}$ 和 $\{\sigma(z_k)\}$ 的有界性。由于涉及映射 A 和 σ （如定义3.4(a)所述）的情形，其证明遵循与 [9, 引理3中2], 相同的步骤，我们省略其证明，下面分析 σ 为条件强制的情形。

命题 4.3 假设 (A_0) 成立，且 $z_0 \neq 0$。固定 $\tilde{r}$ 作为 $\{r_k\}$ 的上界。

(a) 若 σ 如定义 3.4(a) 所述，且 DSG 算法中的 (y_0, c_0) 取为命题 4.2(i) 中所述，则序列 $\{z_k\}$和 $\{\sigma(z_k)\}$ 有界。

(b) 若 σ 是常数为 K_σ 的条件强制函数，且 (y_0, c_0) 在 DSG 算法中取为命题 4.2(ii) 中所述，则序列 $\{\sigma(z_k)\}$ 有界。此外，若 DSG 中的参数α_0, s_0 选择为满足

$$\alpha_0 s_0 > \frac{M_P - q(y_0, c_0) + \tilde{r}}{K_\sigma\, \sigma(z_0)}, \tag{30}$$

则 $\{z_k\}$ 有界。

证明。(a) 与 [9, 引理 3.2] 类似。下面证明 (b) 中的第一个结论，即 $\{\sigma(z_k)\}$ 的有界性。若算法在第 k_0 次迭代时停止，则序列 $\{\sigma(z_k)\}$ 和 $\{z_k\}$ 有限，因此有界。事实上，在后一种情况下，该序列要么停止（若 $r_{k_0} \leq \epsilon$），要么进入有限的内循环，直到 $r_{k_0} \leq \epsilon$。无论哪种情况，序列 $\{z_k\}$ 和 $\{\sigma(z_k)\}$ 都是有限的，因此显然有界。因此，只需假设每次 $k \geq 0$ 都访问步骤 2，从而对每个 $k \geq 0$，$z_k \neq 0$。记 $a_0 := \alpha_0 s_0 \sigma(z_0) > 0$。由 (25) 可推得，对所有$k \geq 1$,

$$c_k - c_0 = \sum_{l=0}^{k-1}(\alpha_l + 1)s_l\sigma(z_l) = \sum_{l=0}^{k-1} s_l\sigma(z_l) + \sum_{l=0}^{k-1}\alpha_l s_l \sigma(z_l)$$

$$\geq \sum_{l=0}^{k-1} s_l\sigma(z_l) + \alpha_0 s_0\sigma(z_0) \geq \sum_{l=0}^{k-1} s_l\|Az_l\| + \alpha_0 s_0\sigma(z_0)$$

$$\geq \|y_k - y_0\| + \alpha_0 s_0\sigma(z_0) = \|y_k - y_0\| + a_0,$$

其中第二个不等式中用到了（$\mathbf{A_0}$），最后一个不等式中用到了（25），最后一个等式中用到了 a_0 的定义。重新排列该式可得

$$c_k - c_0 - \|y_k - y_0\| \geq a_0. \tag{31}$$

根据命题 3.3(i)，我们知道 $(-A(z_k)\ \sigma(z_k)) \in \partial_{r_k} q(y_k,\ c_k)$ 。利用次梯度不等式，对任意 k,

$$\begin{aligned} -\infty < q_0 = q(y_0, c_0) \ &\leq\ q(y_k, c_k) + \langle -A(z_k), y_0 - y_k\rangle + (c_0 - c_k)\sigma(z_k) + r_k \\ &\leq\ q_k + \|A(z_k)\|\|y_k - y_0\| + (c_0 - c_k)\sigma(z_k) + r_k \\ &\leq\ q_k + \sigma(z_k)\left(\|y_k - y_0\| + c_0 - c_k\right) + \tilde{r} \\ &\leq\ q_k - a_0\sigma(z_k) + \tilde{r} \leq q_k + \tilde{r}, \end{aligned}$$

这里我们使用了柯西-施瓦茨不等式，**(A$_0$)**，以及 (31)。上述表达式给出了 $\{\sigma(z_k)\}$有界的条件。事实上，它可以重新整理为

$$\sigma(z_k) \leq \frac{M_D - q_0 + \tilde{r}}{a_0} := b.$$

因此 $\sigma(z_k) \leq b$ 对所有 k 成立，第一条结论的证明至此完成。现在证明，若 (30) 成立，则 $\{z_k\}$也是有界的。事实上，(30) 直接意味着$a_0 = \alpha_0 s_0 \sigma(z_0) > \frac{M_D - q_0 + \tilde{r}}{K_\sigma}$，因此上述表达式变为

$$\sigma(z_k) \leq \frac{M_D - q_0 + \tilde{r}}{a_0} < K_\sigma,$$

这意味着，根据 K_σ 的定义，$\{z_k\}$ 有界。证明完毕。 □

接下来我们证明：如果偏转次梯度算法生成的某次迭代是对偶解，那么偏转次梯度算法的精确版本必定在当前迭代或下一次迭代时停止。对于 σ 的任一类型，此结论均成立。

命题 4.4 假设 (A$_0$) 成立，且对于所有 k，偏转次梯度算法有 $r_k = 0$ 。如果第 k 次偏转次梯度算法迭代是一个对偶解，那么要么 $z_k = 0$ ，要么 $z_{k+1} = 0$。因此，在这种情况下，偏转次梯度算法将在第 k 次或第 $k+1$ 次迭代时停止。

证明。假设在第 k 次迭代时，我们有 $(y_k, c_k) \in S(D)$。这意味着 $q_k = q(y_k, c_k) = M_D$。只需证明，若 $z_k \neq 0$，则 $z_{k+1} = 0$。假设 $z_k \neq 0$，根据 (A$_0$)，显然有

$$\sigma(z_k)\sigma(z_{k+1}) - \|Az_k\|\,\|Az_{k+1}\| \geq 0. \tag{32}$$

取 $(x_{k+1}, z_{k+1}) \in X(y_{k+1}, c_{k+1})$。按照偏转次梯度算法的记号，记 $\varepsilon_k := \alpha_k s_k$。利用 $(y_k, c_k) \in S(D)$ 这一事实以及 q 和 (x_{k+1}, z_{k+1}) 的定义，我们可以写出

$$\begin{aligned} M_D \geq q_{k+1} \ &=\ f(x_{k+1}, z_{k+1}) - \langle Az_{k+1}, y_{k+1}\rangle + c_{k+1}\sigma(z_{k+1}) \\ &=\ f(x_{k+1}, z_{k+1}) - \langle Az_{k+1}, y_k - s_k Az_k\rangle + (c_k + (s_k + \varepsilon_k)\sigma(z_k))\sigma(z_{k+1}) \\ &=\ f(x_{k+1}, z_{k+1}) - \langle Az_{k+1}, y_k\rangle + s_k\langle Az_{k+1}, Az_k\rangle + (c_k + (s_k + \varepsilon_k)\sigma(z_k))\sigma(z_{k+1}) \\ &\geq\ [f(x_{k+1}, z_{k+1}) - \langle Az_{k+1}, y_k\rangle + c_k\sigma(z_{k+1})] + s_k\left(\sigma(z_k)\sigma(z_{k+1}) - \|Az_k\|\,\|Az_{k+1}\|\right) \\ &\quad + \varepsilon_k\sigma(z_k)\sigma(z_{k+1}) \geq q_k + \varepsilon_k\sigma(z_k)\sigma(z_{k+1}) = M_D + \varepsilon_k\sigma(z_k)\sigma(z_{k+1}), \end{aligned}$$

在第三个等式中，我们使用了 DSG 的定义。在倒数第二个不等式中，我们还使用了（32）以及q_k 的定义。这表明 $\varepsilon_k\sigma(z_k)\sigma(z_{k+1}) \leq 0$。由于假设 ε 和 $\sigma(z_k)$ 均为正，因此必有 $\sigma(z_{k+1}) = 0$ ，从而 $z_{k+1} = 0$。 □

下面的定理指出，偏转次梯度算法保证对偶函数单调递增。若初始迭代点取为定理 3.2中的形式，我们知道，对于任一种 σ （强制的或条件强制的）情形，算法都是定义良好的。假设这一点成立，下面结果的证明与 [9, 定理 3.1] 中的步骤类似，因此略去。

定理 4.1 假设偏转次梯度算法生成一个无限序列 $\{(y_k, c_k)\}$，并且对于每个 k，(y_k, c_k) 都不是对偶解。那么 $q(y_{k+1}, c_{k+1}) > q(y_k, c_k)$。

证明。类似于 [9,定理 3。1]。 □

从现在起，我们假设对所有 k，$z_k \neq 0$。换言之，我们假设该方法生成一个无限序列。我们还假设初始迭代点和参数的选取使得，对于两种类型的 σ，先前的结果和性质均成立。下面的技术结果的证明与 [9, 引理 3.3] 中的证明类似，因此略去。

Le理 4.1 由 DSG 算法生成的序列 $\{(x_k, z_k)\}, \{(y_k, c_k)\}$。

(a) 对所有 $k \geq 1$，以下估计均成立

$$f(x_k, z_k) - \langle A(z_k), y_0 \rangle \leq q_k + r_k, \ and \tag{33}$$

$$\sigma(z_k) \sum_{j=0}^{k-1} \alpha_j s_j \sigma(z_j) \leq q_k - q_0 + r_k. \tag{34}$$

(b) 假设对偶解集 $S(D)$ 非空。若 $(\bar{y}, \bar{c}) \in S(D)$，则对所有 k，

$$\|y_{k+1} - \bar{y}\|^2 \leq \|y_k - \bar{y}\|^2 + 2s_k\sigma(z_k)\left[\frac{s_k\sigma(z_k)}{2} + \frac{q_k - \bar{q} + r_k}{\sigma(z_k)} + \bar{c} - c_k\right]. \tag{35}$$

证明。与 [9,引理 3.3] 类似。 □

以下结果适用于 σ 的任一类型。证明中唯一涉及的新结果是：对于 σ 条件强制的情形，强对偶性成立。同样，由于证明技巧与 [9, 引理 3.4]，相似，这里省略证明。

引理 4.2 若序列 $\{z_k\}$ 弱收敛于 0，则 $\{q_k\}$ 收敛到 $\bar{q}$，原始序列 $\{x_k\}$ 有界，且其所有弱聚点都是原始解。

证明。与 [9,引理 3.4] 类似。 □

4.2 DSG-1算法

本节考虑与 [9, 算法 1]，中给出的步长类似的步长，并将其用于我们的特定方案。区别在于选择步长时使用了函数 A（见下文 η_k 的定义）。取两个参数 $\beta > \eta > 0$。我们考虑步长

$$s_k \in [\eta_k, \beta_k], \tag{36}$$

其中 $\eta_k := \min\{\eta, \|A(z_k)\| + \|z_k\|\}$ 且 $\beta_k := \max\{\beta, \sigma(z_k) + \|z_k\|\}$。采用这种 s_k 选择时，我们将 DSG 算法记为 DSG-1。

注4.5 若 (A_0) 成立，则

$$\eta_k \leq \|A(z_k)\| + \|z_k\| \leq \sigma(z_k) + \|z_k\| \leq \beta_k,$$

其中，第一个和最后一个不等式使用了 η_k、β_k 的定义。第二个不等式由 (A_0) 得出。注意，所有迭代均可采用恒定步长。

下一个定理只要求一个满足以下性质的 σ：

$$\text{if } \sigma(w_k) \downarrow 0 \text{ then } \{w_k\} \text{ bounded.} \tag{37}$$

其证明根据对偶序列 $(y_k,\ c_k)$ 是否有界，分别考虑两种情况。对于序列无界的情况，其证明与 [9, 定理 3。2], 中的证明类似，因此略去。对偶序列 $(y_k,\ c_k)$ 有界时，情况略有不同，这是因为我们的步长类型不同，因此在此给出证明。

定理 4.2 假设 σ 是一个增广函数，满足：若 $\sigma(w_k) \downarrow 0$，则 $\{w_k\}$有界的；并假设 $M_P = M_D$。考虑由 DSG-1 生成的原始序列 $\{x_k\}$ 。取参数序列 $\{\alpha_k\}$ ，使得对所有 k 及某个 $\bar{\alpha} > 0$，均有 $\alpha_k \geq \bar{\alpha}$ 。则 $\{x_k\}$ 有界，其所有弱聚点均为原始解，且 $\{q_k\}$ 收敛到最优值 M_P。

证明。取由 DSG-1 生成的对偶序列 $\{(y_k, c_k)\}$。如果 $\{(y_k, c_k)\}$ 无界，则证明与定理 3.2]中相应部分类似。下面考虑 $\{(y_k, c_k)\}$ 有界的情形。由命题 4.1（i），$\sum_k s_k\sigma(z_k) < \infty$。特别地，$\{s_k\sigma(z_k)\}$收敛于 0。另一方面，由 $s_k \geq \min\{\eta, \|A(z_k)\| + \|z_k\|\}$，我们得到

$$s_k\sigma(z_k) \geq \min\{\eta\sigma(z_k), (\|A(z_k)\| + \|z_k\|)\sigma(z_k)\} \geq \min\{\eta\sigma(z_k), \|z_k\|\sigma(z_k)\} > 0,$$

因为 $z_k \neq 0$ 对所有 k成立。由于 $\eta > 0$，我们得出 $\{\|z_k\|\}$ 收敛于 0，或 $\sigma(z_k)$ 收敛于 0。我们将证明这两种情形均蕴含 $\{z_k\}$ 弱收敛于 0。如果 $\{\|z_k\|\}$ 收敛于 0，则 $\{z_k\}$ 强收敛于 0，因而弱收敛于 0。另一方面，如果 $\sigma(z_k)$ 收敛于 0，则根据假设，$\{z_k\}$ 有界。于是，存在一个子序列 $\{z_{k_j}\}$ 弱收敛于某个 $\tilde{z}$。由 $\sigma(\cdot)$ 的弱下半连续性，我们有 $0 \leq \sigma(\tilde{z}) \leq \liminf_{k\to\infty} \sigma(z_{k_j}) = 0$。因此，$\sigma(\tilde{z}) = 0$。$\sigma$ 的性质现在意味着 $\tilde{z} = 0$。因此，整个序列 $\{z_k\}$弱收敛于 0。于是，在 $\{(y_k, c_k)\}$ 有界的情形下，结论由引理 4.2 和零对偶间隙性质 $\bar{q} = M_P$ 得出。□

以下推论成立，因为条件强制的 σ 蕴含强对偶性，同时也验证了（37）。

推论 4.1 若 σ 满足定义*3.4*（b），且满足（a）或（a'），则定理*4.2* 的结论成立。

证明。由于条件强制的 σ 满足（37），强制的 σ也满足该条件。根据定理3.1，强对偶性成立。因此，我们满足定理 4.2的条件。□

上述定理 4.2 给出了 DSG-1 的原始收敛结果，下面的定理给出一个对偶收敛结果；其证明与 [9, 定理 3.3] 相同，故略去。

定理 4.3 若 DSG-1 生成一个无限序列 $\{(y_k, c_k)\}$，则 $\{(y_k, c_k)\}$ 的每个弱聚点（若存在）都是一个对偶解。

证明。参见 [9,定理 3.3]。 □

我们知道，当 $\{c_k\}$ 有界时，$\{y_k\}$ 根据命题4.1有界。反之则不一定成立；只有在额外假设下才成立，该假设要求序列 $\{r_k\}$和 $\sigma(z_k)$以相似速率递减。

(Γ_0)：存在 $R > 0$ 使得对所有 k，都有 $r_k \leq R\sigma(z_k)$，也就是说 $r_k \approx O(\sigma(z_k))$。

下一个结果的证明与 [9, 引理 35] 类似，因此略去。

引理 4.3 假设 **(A_0)** 和 **(Γ_0)** 成立。如果对偶解集非空且 $\{y_k\}$有界的，那么 $\{c_k\}$ 也是有界的。

证明。与 [9,引理 3.5]. □

注 4.6 引理 4.3 在偏转次梯度算法的一般框架下，满足假设 $(\mathbf{A_0})$ 和 $(\mathbf{\Gamma_0})$ 时成立，与步长 s_k 的选择无关。

以下结果将 [9, 命题 3.3] 推广到我们的广义情形。由于我们的函数 σ 和步长不同，因此需要稍作调整的证明。

命题 4.5 假设 $(\mathbf{\Gamma_0})$ 成立，且满足命题 *4.3*(a) 或 (b) 的条件。还假设 DSG-1 生成了一个无限的对偶序列 $\{(y_k, c_k)\}$。若对偶最优集非空，则 $\{(y_k, c_k)\}$ 有界。

证明。在命题 4.3(a) 或 (b) 的条件下，我们有 $\{\sigma(z_k)\}$ 和 $\{z_k\}$ 有界，因此取 $b > 0$ 使得对所有 k，都有 $\sigma(z_k)+\|z_k\| < b$。根据定义，$s_k \le \beta_k \le \max\{\beta, b\} =: b$。ˆ¯特别地，对所有 k，都有 $s_k\sigma(z_k) \le bb = b$。令 R 如 $(\mathbf{\Gamma_0})$ 中所述，并取 $(\bar{y}, \bar{c}) \in S(D)$。我们断言，$\{(y_k, c_k)\}$ 有界。首先证明 $\{c_k\}$ 有界。反设 $\{c_k\}$ 无界。因此存在 k_0，使得对所有 $k \ge k_0$，都有 $c_k \ge M :$ $= \frac{\bar{b}}{2} + R + \bar{c}$。注意到 $q_k \le \bar{q}$，并利用 (35) 中的估计，我们得到

$$\begin{aligned}
\|y_{k+1} - \bar{y}\|^2 &\le \|y_k - \bar{y}\|^2 + 2s_k\sigma(z_k)\left[\frac{s_k\sigma(z_k)}{2} + \frac{q_k - \bar{q} + r_k}{\sigma(z_k)} + \bar{c} - c_k\right] \\
&\le \|y_k - \bar{y}\|^2 + 2s_k\sigma(z_k)\left[\frac{\bar{b}}{2} + \frac{r_k}{\sigma(z_k)} + \bar{c} - c_k\right] \\
&\le \|y_k - \bar{y}\|^2 + 2s_k\sigma(z_k)\left[\frac{\bar{b}}{2} + R + \bar{c} - c_k\right] \\
&\le \|y_k - \bar{y}\|^2
\end{aligned} \tag{38}$$

对所有 $k \ge k_0$。由此可知，$\{\|y_k - \bar{y}\|\}$ 是一个递减序列，因此 $\{y_k\}$ 有界。根据引理4.3，这将导致矛盾。因此，对偶序列有界。 □

定理 4.4将在下文证明，它表明由 DSG-1 生成的整个对偶序列强收敛到一个对偶解。 [9] 中的定理 3.4 依赖于 Fej´er 收敛性质，只能证明对偶序列弱收敛。我们的证明受 [10] 中的定理 5.1 启发，并利用 q的性质。

定理 4.4 若 $(\mathbf{A_0})$ 成立，且参数序列 $\{\alpha_k\}$ 满足：对所有 k，$\alpha_k \ge \bar{\alpha}$，并且某个 $\bar{\alpha} > 0$，则下列结论成立。

(i) 假设 σ 是一个满足 (37)的增广函数，并假设 $M_P = M_D$。若由 DSG-1 生成的对偶序列有界，则 $S(D) \ne \varnothing$，且对偶序列强收敛到一个对偶解。(ii) 假设 $(\mathbf{\Gamma_0})$ 成立，且满足命题 *4.3*(a) 或 (b) 的条件。若 $S(D) \neq \varnothing$，则由 DSG-1 生成的对偶序列强收敛到某个对偶解。

证明. (i) 由于对偶序列 $\{(y_k, c_k)\}$ 有界，因此由命题 4.1(a)(c)，它强收敛到某个 $(\tilde{y}, \tilde{c})$。我们现在只需证明极限 $(\tilde{y}, \tilde{c}) \in S(D)$。事实上，我们可以写成

$$M_P = \lim_k q(y_k, c_k) \le q(\tilde{y}, \tilde{c}) \le M_P, \tag{39}$$

其中等式由定理4.2 和强对偶性（见定理 3.1）可得。注意，由于序列 $\{q_k\}$ 是递增的，因此我们可以写成极限。第一个不等式由以下事实得出：

q 是弱上半连续的（因此也是强上半连续的），最后一个不等式则由以下事实得出：$q(\tilde{y},\tilde{c})$ $\leq M_D \leq M_P$。因此，我们证明了$(\tilde{y},\tilde{c}) \in S(D)$，(i) 的证明至此完成。现在证明(ii)。由于对偶集非空，根据命题 4.5，对偶序列$\{(y_k,c_k)\}$是有界的。由(i)，我们有 $\{(y_k,c_k)\}$ 强收敛到某个$(\tilde{y},\tilde{c}) \in S(D)$。□

下面这个直接推论刻画了对偶解的存在性。

推论 4.2 若 $(\mathbf{A_0})$和 $(\mathbf{\Gamma_0})$成立，且满足命题 *4.3*(a) 或 (b) 的条件。假设 $\alpha_k \geq \bar{\alpha} > 0$ 对所有 k均成立。以下陈述等价：

(a) 由 DSG-1 生成的对偶序列有界。

(b) 对偶集非空。

证明。该证明直接源于以下事实：这些假设确保我们满足定理 4.4中 (i) 和 (ii) 两部分的条件。□

4.3 算法 DSG-2

本节采用 [9, 算法 2], 中提出的相同步长，该步长可确保偏转次梯度算法在有限步内收敛。本节将说明，使用映射 A 构造拉格朗日函数时，这些收敛结果仍然成立。

取 $\beta > 0$ 和一个序列 $\{\theta_k\} \subset \mathbb{R}_+$，使得 $\sum_j \theta_j = \infty$，且对所有 k均有 $\theta_k \leq \beta$。考虑步长

$$s_k \in [\eta_k, \beta_k], \tag{40}$$

其中 $\eta_k := \theta_k/\sigma(z_k)$ 且 $\beta_k := \beta/\sigma(z_k)$。采用这种步长选取方式的偏转次梯度算法记为 DSG-2。

下面的结果推广了[9, 定理 3。5]仔细考察 [9, 定理 3。5], 中的证明可以发现，只需假设 σ 满足（37），便可弱化这些假设。由于证明与 [9, 定理 3。5], 中的证明类似，此处从略。

定理 4.5 设 $\{(x_k,\ z_k)\}$ 和 $\{(y_k,\ c_k)\}$ 是由 DSG-2生成的序列。设参数序列 $\alpha_k \geq \bar{\alpha} > 0$。假设 σ 满足 (37)。那么以下情况恰有一种成立：

(a) 存在一个 k，使得 DSG-2 在第 k 次迭代时停止。因此，$x_{\bar{k}}$ 和（$y_{\bar{k}}$，$c_{\bar{k}}$）分别是ϵ-最优原始解和 ϵ-最优对偶解。在这种情况下，$\{(y_k,c_k)\}$ 必定有界。

(b) 对偶序列 $\{(y_k,c_k)\}$ 无界。在这种情况下，$\{z_k\}$ 弱收敛到 0，$\{q_k\}$收敛到 M_P，原始序列 $\{x_k\}$ 有界，且其所有弱聚点都是原始解。

证明。与 [9,定理 3.5] 类似。□ 由此直接得到以下结果。

推论 4.3 设 $\{(x_k,z_k)\}$ 和 $\{(y_k,c_k)\}$ 是由 DSG-2生成的序列。假设参数序列 $\alpha_k \geq \bar{\alpha} > 0$。设 σ 是强制的或条件强制的。则定理 *4.5*的结论成立。

证明。该结论由以下事实直接得出：强制性和条件强制性均蕴含条件（37）。□

5 致谢

感谢 C. Yalcin Kaya 就 DSG 算法的各个方面进行了有益的讨论。

6 结论

我们对无限维空间中的原始-对偶问题进行了详细分析，其中对偶问题是通过使用一个拉格朗日函数得到的，该函数在线性项中包含一个映射 A （见定义（5）3.5），以及一个在某些情况下不必是强制的增广函数（见定义3.4(a’)）。对于这类原始-对偶对，我们证明了强对偶性。此外，我们表明，不精确 DSG 方法具有与 A 为恒等映射时相同的原始和对偶收敛性质。具体而言，我们证明，原始序列的每个聚点都是原始解，并且在某些技术性假设下，对偶序列强收敛到对偶解（见定理 4.4）。

我们的分析为将偏转次梯度算法应用于具有挑战性的最优控制问题开辟了道路，这类问题属于无限维优化问题。尤其值得将本文的结果应用于纯罚函数法，即 $A = 0$。后者的情形不在 [15] 的结果涵盖范围内。

参考文献

[1] D.P. Bertsekas，《非线性规划》，Athena Scientific，第二版，1999年。[2] H. Brezis，《泛函分析、Sobole夫空间与偏微分方程》，Springer，柏林，2011年。[3] R.S. Burachik，关于增广拉格朗日对偶的原始收敛，《优化》，60：979–990（2011年）。[4] R.S. Burachik，关于非光滑优化的渐近拉格朗日对偶，《ANZIAM jour- nal》，58：93–123（2017年）。[5] R.S. Burachik、R.N. Gasimov、N.A. Ismayilova 和 C.Y. Kaya，基于尖锐增广拉格朗日函数的对偶问题修正次梯度算法，《J. Global Optim》，34：55–78（2006年）。[6] R.S. Burachik 和 A.N. Iusem，《集值映射与单调算子的扩张》，Springer，柏林，2008年。[7] R.S. Burachik、A.N. Iusem 和 J.G. Melo，一种具有尖锐拉格朗日函数的原始-对偶修正次梯度算法，《J. Global Optim》，46：55–78（2010年）。[8] R.S. Burachik、A.N. Iusem 和 J.G Melo，一般增广拉格朗日函数的对偶与精确罚，《J. Optim. Theory Appl》，147(1)：125–140（2010年）。[9] R.S. Burachik、A.N. Iusem 和 J.G. Melo，基于增广拉格朗日函数的原始–对偶问题非精确修正次梯度算法，《J. Optim. Theory Appl》，157：108–131（2013年）。[10] R.S. Burachik、A.N. Iusem 和 J.G. Melo，非光滑与非凸优化的精确罚映射，《Optim.》，64(4)：717–738（2015年）。[11] R.S. Burachik、W.P. Freire 和 C.Y. Kaya，基于广义增广拉格朗日对偶的非光滑与非凸优化内部上图方向法，《J. Global Optim.》：60 (3)，501–529（2014年）。[12] R.S. Burachik 和 C.Y. Kaya，罚函数法的更新规则与收敛结果，《J. Ind. Manage. Optim.》，3(2)：381–403（2007年）。

[13] R.S. Burachik 和 C.Y. Kaya，采用一般增广拉格朗日对偶的偏转次梯度方法及其对罚函数方法的启示。载于：R.S. Burachik、Yao、J.C.（编），《优化与控制中的变分分析和广义微分》，Springer 优化及其应用，47：109–132。Springer，纽约，（2010年）。[14] R.S. Burachik 和 C.Y. Kaya，一种用于非凸优化且带有罚参数更新的增广罚函数方法，《Nonlinear Anal. Theory Methods Appl.》，75(3)：1158–1167（2012年）。[15] R.S. Burachik、C.Y. Kaya 和 M. Mammadov，一种用于非凸优化的非精确修正次梯度算法，《Comput. Optim. Appl》，45，1–24（2010年）。[16]R.S. Burachik、C.Y. Kaya 和 C.J. Price，一种基于舍入加权-ℓ_1 拉格朗日对偶的原始-对偶罚函数方法，《Optim.》（2021年）。[17] R.S. Burachik 和 A.M. Rubinov，抽象凸性与增广拉格朗日函数，《SIAM J. Optim.》，18：413-436（2007年）。[18] R.S. Burachik 和 X.Q. Yang，渐近强对偶，《Numer. Algebra, Control Optim.》，1(3)：539–548（2011年）。[19] R.S. Burachik、X.Q. Yang 和 Y.Y. Zhou，半无限规划问题中增广拉格朗日乘子的存在性，《J. Optim. Theory Appl》，173：471–503（2017年）。[20] Y. Chen 和 M. Chen，非线性规划的扩展对偶，《Comput. Optim. Appl.》，47(1)：33–59（2010年）。[21] M.V. Dolgopolik，锥约束优化的增广拉格朗日函数：全局鞍点的存在性与精确罚函数性质，《J. Global Optim.》，71：237–296（2018年）。[22] M.V. Dolgopolik，罚函数与增广拉格朗日函数全局精确性的统一方法 I：参数精确性，《J Optim. Theory Appl.》，176：728–744（2018年）。[23] R.N. Gasimov，非凸规划中的增广拉格朗日对偶与非光滑优化方法，《J. Global Optim.》，24：187–203（2002年）。[24] X.X. Huang 和 X. Q. Yang，一种统一的增广拉格朗日方法：对偶与精确罚，《Math. Oper. Res》，28：533–552（2003年）。[25] X.X. Huang 和 X. Q. Yang，增广拉格朗日对偶理论的进一步研究，《J. Global Optim.》，31：193–210（2005年）。[26] E. Kreyszig，《泛函分析及其应用导论》，Wiley，1978年。[27] D. Li，非线性非凸优化中的鞍点生成，《Nonlinear Anal.》，30：4339–4344（1997年）。[28] A. Nedic 和 A. Ozdaglar，利用增广拉格朗日函数研究非凸优化对偶的几何框架，《J. Global Optim.》，40：545–573（2008年）。[29] Y. Nesterov，《凸优化导论：基础课程》，Kluwer Academic Publishers，2004年。[30] B. Polyak，《优化导论》，Optimization Software 公司，1987年。[31] R.T. Rockafellar 和 R.J.B. Wets，《变分分析》，Springer 出版社，柏林，1998年。

[32] A.M. Rubinov, 《抽象凸性与全局优化》, Kluwer Acadernic Publishers, 2000年。[33] A.M. Rubinov 和 X. Yang, 《约束非凸优化中的拉格朗日型函数》, 施普林格科学与商业媒体, 2013年。[34] C.Y. Wang、X.Q. Yang 和 X.M. Yang, 《非线性增广拉格朗日函数与对偶理论》, Math. Oper. Res, 38：740–760（2012年）。[35] L. Zhang 和 X. Yang, 《非线性规划中结合变量变换的增广拉格朗日方法》, Nonlinear Anal., 69：2095–2113（2008年）。[36] Y.Y. Zhou 和 X.Q. Yang, 《关于对偶与精确罚的若干结果》, J. Global Optim., 29：497–509（2004年）。[37] Y.Y. Zhou 和 X.Q. Yang, 《增广拉格朗日函数、非二次增长条件与精确罚》, Oper. Res. Lett., 34：127–134（2006年）。[38] Y.Y. Zhou 和 X.Q. Yang, 《通过增广拉格朗日函数实现优化中的对偶与罚函数法及其应用》, J. Optim. Theory Appl., 140：171–188（2009年）。[39] Y.Y. Zhou 和 X.Q. Yang, 《约束优化问题的增广拉格朗日函数》, J. Global Optim., 52：95–108（2012年）。