

L_1 正则化问题的对偶性理论*

吴焚供^{1,2}

(1. 华南师范大学数学科学学院, 广州 广东 510631;
2. 广东第二师范学院数学系, 广州 广东 510303)

摘要: L_1 正则化问题是一个非光滑的无约束最优化问题, 在变量选择, 数据压缩和图像处理等领域有广泛的应用。给出了 L_1 问题最优解存在的新的必要条件和充分条件, 利用这些条件构造出 L_1 正则化问题的一个 Mond-Weir 型对偶问题, 最后给出了相应的弱对偶定理和强对偶定理。

关键词: L_1 正则化; 最优解; 对偶问题

中图分类号: O224 文献标志码: A 文章编号: 0529-6579 (2015) 01-0010-03

Duality Theorem for L_1 -Regularization Problem

WU Fengong^{1,2}

(1. Department of Mathematics, South China Normal University, Guangzhou 510631, China;
2. Department of Mathematics, Guangdong University of Education, Guangzhou 510303, China)

Abstract: L_1 -regularization problem is a non-smooth unconstrained optimization problem, which is widely used in the fields such as variable selection, data compression and image processing. Optimality conditions for the solution of L_1 -regularization problem is given. And a Mond-Weir type dual problem for L_1 -regularization problem is formulated, by using these optimal conditions. Finally a weak duality theorem and a strong duality theorem are proved.

Key words: L_1 -regularization; optimality condition; dual problem

正则化问题近年来备受关注, 许多研究者考虑如下的 L_p 最小化问题

$$\min F(x) = f(x) + \rho \|x\|_p^p$$

其中 $f(x): \mathbf{R}^n \rightarrow \mathbf{R}$ 为一光滑函数, ρ 为一给定的非负正则化参数, $p \in [0, 1]$, 变量 $x \in \mathbf{R}^n$, $\|x\|_p$ 为定义在 \mathbf{R}^n 上的 L_p 拟范数, 当 $0 < p \leq 1$ 时, 定义

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}$$

当 $p = 0$ 时, 定义 $\|x\|_0 = |\{i | x_i \neq 0, i = 1, 2, \dots, n\}|$, 即 x 中非零元的个数。 L_p 问题在地球物理学, 变量选择, 数据压缩, 图像处理, 感知网络及计算机视觉等领域有广泛的应用, 具体可参考文 [1-7]。

L_p 正则化问题的一个重要情况是如下的 $L_2 -$

L_p 最小化问题

$$\min F(x) = \|Ax - b\|^2 + \rho \|x\|_p^p \quad (1)$$

其中 $\|\cdot\|$ 为欧几里得范数, $A \in \mathbf{R}^{m \times n}$, $m < n$ 甚至 $m \ll n$, $b \in \mathbf{R}^m$ 。特别当 $p = 0$ 时, 称为 $L_2 - L_0$ 问题, 该问题旨在找出 $Ax = b$ 的最稀疏解 (换言之, 也就是所有解中非零元最少者), 统计学中的变量选择方法便可以视为 $L_2 - L_0$ 问题。然而, L_0 正则化问题需要求解一个困难的组合优化问题, Tibshirani^[7] 和 Donoho^[1] 等分别从变量选择和信号的稀疏重建的角度提出了 L_1 正则化

$$\min F(x) = f(x) + \rho \|x\|_1 \quad (2)$$

该问题在文 [7] 和文 [1] 中分别被称为 LASSO 和 Basis Pursuit。他们证明了 L_1 正则化问题

* 收稿日期: 2014-03-07

基金项目: 广东省教育厅科研项目“育苗工程”(自然科学)资助项目(2013LYM_0061)

作者简介: 吴焚供(1980年生), 男; 研究方向: 最优化算法理论; E-mail: wufengong@gdei.edu.cn

与 L_0 正则化问题在稀疏重建的意义下是等价的, 从而 L_0 正则化所要求的 NP 组合优化问题可以转化为 L_1 罚优化问题。基于上述学者的工作, L_1 正则化成为当今数据分析的主流工具之一。

L_1 正则化问题是 Lipschitz 连续且在 $\mathbf{R}^n \setminus X^0$ 上连续可微的, 其中 $X^0 = \{x | x_i = 0, i \in N\}, N = \{1, 2, \dots, n\}$ 。实际上, 当 $f(x)$ 连续可微时, 式 (2) 中的 $F(x)$ 按照 Clark^[8] 定义的广义导数是处处可微的。特别当 $f(x)$ 是凸函数时, 根据文 [8], L_1 问题最优解存在的充要条件可叙述为

$$\begin{cases} \nabla_i f(x) + \rho \text{sign}(x_i) = 0, & |x_i| > 0, \\ |\nabla_i f(x)| \leq \rho, & x_i = 0 \end{cases} \quad (3)$$

其中 $\nabla_i f(x) = \partial f(x) / \partial x_i$, $\text{sign}(t)$ 为符号函数。条件 (3) 在文 [9] 中被叙述为如下的等价形式在这些理论的基础上

$$x = \text{sign}(x - \tau \nabla f(x)) \odot \max\{|x - \tau \nabla f(x)| - \tau \rho, 0\}, \tau > 0$$

或

$$\max\{\nabla f(x) - \rho, \min\{x, \nabla f(x) + \rho\}\} = 0$$

在这些理论的基础上, 一些有效求解 L_1 问题的算法得以建立, 具体可参考文 [9-11] 等。

然而 L_1 正则化问题毕竟是非光滑优化问题, 若能构造出其光滑的对偶问题, 则更有效的求解该问题便是有可能的。另一方面, 早在 1974 年 Mond^[12] 和 1981 年 Mond 和 Weir^[13] 便考虑了一类特殊的不可微优化问题的对偶问题, 受其工作的启发, 运用类似的方法构造出 L_1 正则化问题的一个光滑的对偶问题, 我们称之为 Mond-Weir 型对偶。

1 最优性条件

考虑原始问题 (P)

$$\min F(x) = f(x) + \rho \|x\|_1$$

令 $x_0 = (x_1^0, x_2^0, \dots, x_n^0) \in \mathbf{R}^n$, 记

$$Z_0 = \{z \mid z^T \nabla f(x_0) + \rho \sum_{i \in \beta(x_0)} \text{sign} x_i^0 z_i + \rho \sum_{i \in \overline{\beta(x_0)}} |z_i| < 0\}$$

其中, $\beta(x_0) = \{i \mid x_i^0 = 0, i \in N\}$, $\overline{\beta(x_0)} = \{i \mid x_i^0 \neq 0, i \in N\}$, $N = \{1, 2, \dots, n\}$ 。

引理 1 若对任意的 $z \in \mathbf{R}^n$, $Z_0 = \emptyset$, 则存在 $w \in \mathbf{R}^n$, 使得 $|w_i| \leq 1, (i = 1, 2, \dots, n)$, $\|x_0\|_1 = x_0^T w$, 且 $\nabla f(x_0) + \rho w = 0$ 。

证明 记 $S = \{y \mid y = \nabla f(x_0) + \rho w, w^T x_0 = \|x_0\|_1, w \in \mathbf{R}^n, |w_i| \leq 1, i = 1, 2, \dots, n\}$, 若不存在 $w \in \mathbf{R}^n$ 使命题成立, 则 $0 \notin S$, 由于 S 是闭凸集, 根据凸集分离定理, 存在 $z' \in \mathbf{R}^n$ 及 $\alpha \in \mathbf{R}$,

使得

$$z'^T y \leq \alpha < 0, \forall y \in S \quad (4)$$

取 $w \in \mathbf{R}^n$, 使得

$$w_i = \begin{cases} \text{sign} x_i^0, & i \in \overline{\beta(x_0)}; \\ \text{sign} z'_i, & i \in \beta(x_0) \end{cases}$$

则 $|w_i| \leq 1, i = 1, 2, \dots, n, w^T x_0 = \|x_0\|_1$, 若记 $y = \nabla f(x_0) + \rho w$, 则有 $y \in S$, 所以由式 (4) 有 $z'^T y = z'^T \nabla f(x_0) + \rho \sum_{i \in \beta(x_0)} \text{sign} x_i^0 z'_i + \rho \sum_{i \in \overline{\beta(x_0)}} |z'_i| < 0$ 这与假设 $Z_0 = \emptyset$ 矛盾。

定理 2 (必要条件) 若 x_0 是问题 (P) 的解, 则存在 $w \in \mathbf{R}^n$, 使得

$$\nabla f(x_0) + \rho w = 0 \quad (5)$$

$$|w_i| \leq 1, i = 1, 2, \dots, n \quad (6)$$

$$\|x_0\|_1 = x_0^T w \quad (7)$$

证明 若 x_0 是问题 (P) 的解, 则对任意给定的 $z \in \mathbf{R}^n$, 及任意的实数 $\alpha > 0$, 有

$$\begin{aligned} F(x_0 + \alpha z) - F(x_0) &= \\ \alpha \nabla f(x_0)^T z + \rho \alpha \sum_{i \in \beta(x_0)} \text{sign} x_i^0 z_i + \\ \rho \alpha \sum_{i \in \overline{\beta(x_0)}} |z_i| + o(\alpha) &\geq 0, \end{aligned}$$

两边除以 α 并令 $\alpha \rightarrow 0^+$, 则对任意的 $z \in \mathbf{R}^n$,

$$\nabla f(x_0)^T z + \rho \sum_{i \in \beta(x_0)} \text{sign} x_i^0 z_i + \rho \sum_{i \in \overline{\beta(x_0)}} |z_i| \geq 0,$$

即 $Z_0 = \emptyset$, 所以由引理, 定理结论成立。

定理 3 (充分条件) 若 f 为凸函数, 且存在 $x_0, w \in \mathbf{R}^n$ 使 (x_0, w) 满足 $\nabla f(x_0) + \rho w = 0$ 及 $\|w\|_1 \leq 1$, 则 x_0 是问题 (P) 的解。

证明 对任意的 $x \in \mathbf{R}^n$, 由 $f(x)$ 是凸函数, 所以

$$\begin{aligned} F(x) - F(x_0) &= f(x) + \rho \|x\|_1 - f(x_0) - \rho \|x_0\|_1 \geq \\ & (x - x_0)^T \nabla f(x_0) + \rho \|x\|_1 - \rho \|x_0\|_1 \end{aligned}$$

再由式 (5) - (6) 及 $x^T w \leq \|x\|_1 \cdot \|w\|_1$, 有

$$\begin{aligned} F(x) - F(x_0) &\geq \rho \left(-x^T w + x_0^T w + \right. \\ & \left. \|x\|_1 - \|x_0\|_1 \right) = \rho (\|x\|_1 - x^T w) \geq 0 \end{aligned}$$

2 对偶性定理

以下我们均假设 f 是凸函数, 我们将建立起原问题 (P) 与如下问题 (D) 的对偶关系。

$$\begin{aligned} \text{(D)} \quad \max G(u) &= f(u) - u^T \nabla f(u) \\ \text{s. t. } \nabla f(u) + \rho w &= 0 \end{aligned} \quad (8)$$

$$\|w\|_1 \leq 1 \quad (9)$$

记问题 (P) 的可行域为 Ω_P , 问题 (D) 的可行域为 Ω_D 。

定理 4 (弱对偶定理) 问题 (P) 的下确界大于或等于问题 (D) 的上确界。

证明 对任意 $x \in \Omega_P$, $(u, w) \in \Omega_D$,

$$F(x) - G(u) = f(x) + \rho \|x\|_1 - f(u) + u^T \nabla f(u) \geq (x - u)^T \nabla f(u) + \rho \|x\|_1 + u^T \nabla f(u)$$

根据式 (8) - (9), 有

$$F(x) - G(u) \geq \rho (\|x\|_1 - x^T w) \geq 0$$

所以结论成立。

定理 5 (强对偶定理) 若 $x_0 \in \mathbf{R}^n$ 是原问题 (P) 的一个最优解, 则若存在 $w \in \mathbf{R}^n$, 使 $(x_0, w) \in \Omega_D$, 则 (x_0, w) 必是问题 (D) 的解, 且两问题的最值相等。

证明 若 x_0 是问题 (P) 的解, 则由定理 2 存在 w 使式 (5) - (7) 成立, 而且有

$$G(x_0) = f(x_0) - x_0^T \nabla f(x_0) =$$

$$f(x_0) + \rho x_0^T w = f(x_0) + \rho \|x_0\|_1 = F(x_0)$$

再由定理 4, 只要 $(x_0, w) \in \Omega_D$, 则 $(x_0, w) \in \Omega_D$ 必是 (D) 的解。

参考文献:

- [1] CHEN S S, DONOHO D L, SAUNDERS M A. Atomic decomposition by basis pursuit [J]. *SIAM Journal on Scientific Computing*, 1998, 20(1): 33 - 61.
- [2] CHARTRAND R. Exact reconstruction of sparse signals via nonconvex minimization [J]. *Signal Processing Letters*, IEEE, 2007, 14(10): 707 - 710.
- [3] CHARTRAND R, STANEVA V. Restricted isometry properties and nonconvex compressive sensing [J]. *Inverse Problems*, 2008, 24(3): 035020.

- [4] CHEN X, ZHOU W. Smoothing nonlinear conjugate gradient method for image restoration using nonsmooth nonconvex minimization [J]. *SIAM Journal on Imaging Sciences*, 2010, 3(4): 765 - 790.
- [5] DONOHO D L. Compressed sensing [J]. *Information Theory*, IEEE Transactions on, 2006, 52(4): 1289 - 1306.
- [6] FAN J, LI R. Variable selection via nonconcave penalized likelihood and its oracle properties [J]. *Journal of the American Statistical Association*, 2001, 96(456): 1348 - 1360.
- [7] TIBSHIRANI R. Regression shrinkage and selection via the lasso [J]. *Journal of the Royal Statistical Society: Series B (Methodological)*, 1996: 267 - 288.
- [8] CLARKE F H. Optimization and nonsmooth analysis [M]. SIAM, 1990.
- [9] HALE E T, YIN W, ZHANG Y. Fixed-point continuation for l_1 -minimization; Methodology and convergence [J]. *SIAM Journal on Optimization*, 2008, 19(3): 1107 - 1130.
- [10] SHEVADE S K, KEERTHI S S. A simple and efficient algorithm for gene selection using sparse logistic regression [J]. *Bioinformatics*, 2003, 19(17): 2246 - 2253.
- [11] FU W J. Penalized regressions; the bridge versus the lasso [J]. *Journal of Computational and Graphical Statistics*, 1998, 7(3): 397 - 416.
- [12] BERTRAM M. A class of nondifferentiable mathematical programming problems [J]. *Journal of Mathematical Analysis and Applications*, 1974, 46: 169 - 174.
- [13] MOND B, WEIR T. Generalized concavity and duality [C]. *Generalized Concavity in Optimization and Economics*, 1981: 263 - 279.