

# 原始 -对偶问题

## 1 原始 -对偶问题的发展与应用

### 1.1 理论发展

在线性规划早期发展中最重要发现就是对偶问题，即每一个线性规划问题 (称为原始问题) 都有一个与它对应的对偶线性规划问题 (称为对偶问题)。对偶问题的发展：

- 1928 年美籍匈牙利数学家 J.von 诺伊曼在研究对策论时已发现线性规划与对策论之间存在着密切的联系。两零和对策可表达成线性规划的原始问题和对偶问题
- 1947 年提出对偶理论
- 1951 年 G.B. 丹齐克引用对偶理论求解线性规划的运输问题，研究出确定检验数的位势法原理
- 1954 年 C. 莱姆基提出对偶单纯形法

原始 -对偶方法在对偶理论的基础上逐渐发展起来。原始 -对偶方法的发展：

- 1987 年 Megiddo 提出原始 -对偶算法 ( [9])
- 1992 年 Mehrotra 提出了求解线性规划的适应算法 ( [10])

- 1994 年 Nesterov、Nemirovskii 把解线性规划的相关理论推广到了凸规划、半定规划及二阶锥规划上 ([11])
- 1997 年 Nesterov 和 Todd 把原始对偶算法推广到一个受限制的凸问题上 ([12])

原始 -对偶问题的理论发展比较成熟，并且具有较好的性质、结论，并能应用到很多涉及最值的问题，故其应用越来越广泛。

## 1.2 经济中的应用

对偶理论主要研究经济学中的相互确定关系，涉及到经济学的诸多方面。产出与成本的对偶、效用与支出的对偶，是经济学中典型的对偶关系。经济系统中还有许多其他这样的对偶关系。关于原始 -对偶问题在经济学的应用不再一一赘述。

## 1.3 图像中原始 -对偶问题的应用

在图像去噪的最小化问题中，目标函数不可微，其对应的欧拉 -拉格朗日方程很难求解，像牛顿法之类的高阶方法无法应用，而原始 -对偶方法在处理最小化问题时得到了很好的效果，但其也有缺陷，如数值求解问题、求解原始 -对偶向量的额外花费等。

原始 -对偶问题在图像中的应用是 PDE 图像处理中一个较为崭新的思想，最早是 1999 年 Chan、Golub、Mulet 在《A nonlinear primal-dual method for variation-based image restoration》中考虑图像复原中的经典的最小化问题的对偶问题，自此之后，原始 -对偶问题广泛的与其他方法结合被应用到图像处理中。几个成功的原始 -对偶方法：

- CGM 原始 -对偶方法

- Chambolle 对偶方法

$$\inf_u \frac{1}{2} \int_{\Omega} (Ku - f)^2 dx + \lambda \int_{\Omega} |\Delta u| \quad (1)$$

$$\Leftrightarrow \inf_u \sup_{|p| \leq 1} \frac{1}{2} \int_{\Omega} (Ku - f)^2 dx + \lambda \int_{\Omega} u \operatorname{div} p dx \quad (2)$$

$$\Leftrightarrow \sup_{|p| \leq 1} \inf_u \frac{1}{2} \int_{\Omega} (Ku - f)^2 + \lambda \int_{\Omega} u \operatorname{div} p dx \quad (3)$$

$$\Leftrightarrow \sup_{|p| \leq 1} -\frac{\lambda^2}{2} \int_{\Omega} |K^{-T} \operatorname{div} p - \frac{f}{\lambda}|^2 dx \quad (4)$$

- 半平滑牛顿法
- 原始对偶有效集方法
- 原始对偶混合梯度方法

原始 -对偶问题在解决规划问题时具有较好的性质，所以被广泛应用在图像处理中的很多方面，如：

- 原始 -对偶与 L-1 TV 图像复原结合 ([7])
- 原始 -对偶与小波图像修复结合 ([8])
- 原始 -对偶与 Bregman 迭代结合 ([4])

## 2 对偶空间

为了有助于原始 -对偶问题的理解，首先介绍两个概念：对偶空间、原始 -对偶向量。维基百科 ([1]) 对对偶空间是这样叙述的：对偶空间构造是行向量  $(1 \times n)$  与列向量  $(n \times 1)$  的关系的抽象化。

代数的对偶空间: 设  $V$  为在域  $F$  上的向量空间，定义其对偶空间  $V^*$  为由  $V$  到  $F$  的所有线性函数的集合。即是  $V$  的标量线性变换。 $V^*$  本身是

$F$  的向量空间并且拥有加法及标量乘法：

$$\begin{aligned}(\phi + \varphi)(x) &= \phi(x) + \varphi(x) \\(a\phi)(x) &= a\phi(x)\end{aligned}\tag{5}$$

$\forall \phi, \varphi \in V^*, \forall a \in F, \forall x \in V$ 。可以看出，原始与对偶可能存在一定的互逆变化之意。

### 3 原始 -对偶向量

定义 ([2])：设  $V$  是  $R$  上的有限维向量空间。线性映射  $\omega: V \rightarrow R$  称为  $V$  上的对偶向量。

例：设  $V$  为全体  $2 \times 1$  实矩阵的集合，则它在矩阵加法和数乘规则下构成 2 维向量空间。以  $\omega$  代表任一  $1 \times 2$  实矩阵  $(c, d)$ ，其对  $V$  的任一元素  $v = \begin{pmatrix} a \\ b \end{pmatrix}$  的作用可用矩阵乘法定义： $\omega(v) = (c, d) \begin{pmatrix} a \\ b \end{pmatrix}$ 。这样  $\omega$  即是  $V$  上的对偶向量。

### 4 原始 -对偶问题

例：某厂生产甲乙两种产品，其所需资源和单位价格如下表所示：

资源限量	甲	乙	资源总数
煤	9	4	360
电	4	5	200
油	3	10	300
单位价格	7	12	*

考虑如何决定生产方案，使总收入最大？设生产甲、乙产品的产量分别为  $x_1, x_2$ ：

$$\begin{aligned}
 \max \quad & z = 7x_1 + 12x_2 \\
 & 9x_1 + 4x_2 \leq 360 \\
 s.t. \quad & 4x_1 + 5x_2 \leq 200 \\
 & 3x_1 + 10x_2 \leq 300 \\
 & x_1, x_2 \geq 0
 \end{aligned} \tag{6}$$

假设生产单位不再自己生产产品，而是将全部资源出卖，问购买方如何购买价格，双方才能成交？

设煤、油、电的购买单价为  $y_1, y_2, y_3$ ：

$$\begin{aligned}
 \min \quad & \omega = 360y_1 + 200y_2 + 300y_3 \\
 & 9y_1 + 4y_2 + 3y_3 \geq 7 \\
 s.t. \quad & 4y_1 + 5y_2 + 10y_3 \geq 12 \\
 & y_1, y_2, y_3 \geq 0
 \end{aligned} \tag{7}$$

对偶问题与原始问题之间存在着下列关系：

- 目标函数对原始问题是极大化，对对偶问题则是极小化。
- 原始问题目标函数中的收益系数是对偶问题约束不等式中的右端常数，而原始问题约束不等式中的右端常数则是对偶问题中目标函数的收益系数。
- 原始问题和对偶问题的约束不等式的符号方向相反。
- 原始问题约束不等式系数矩阵转置后即为对偶问题的约束不等式的系数矩阵。
- 原始问题的约束方程数对应于对偶问题的变量数，而原始问题的变量数对应于对偶问题的约束方程数。

- 对偶问题的对偶问题是原始问题，这一性质被称为原始和对偶问题的对称性。

由此可见，原始问题和对偶问题一个是取最大，一个是取最小，并且如果原始问题有最优解，对偶问题也有最优解，且最优值相等。

## 5 论文中原始 -对偶问题

在《A NONLINEAR PRIMAL-DUAL METHOD FOR TOTAL VARIATION-BASED IMAGE RESTORATION》中原始问题:

$$\min_u \int_{\Omega} (\alpha |\nabla u| + \frac{1}{2} (Ku - z)^2) dx dy \quad (8)$$

其对应的欧拉 -拉格朗日方程为:

$$-\alpha \nabla \left( \frac{\nabla u}{|\nabla u|} \right) + K^*(Ku - z) = 0 \quad (9)$$

可以看出  $|\nabla u|$  出现在分母上一般很难线性化，并且可能出现分母等于零的情况。我觉得作者引入  $\omega$  的初衷不是想到了原始 -对偶问题，而是想要除此处的奇异性，用  $\omega$  表示奇异部分  $\frac{\nabla u}{|\nabla|}$ ，如此变换 (9) 转化为:

$$-\alpha \nabla \omega + K^*(Ku - z) = 0 \quad (10)$$

而  $\omega = \frac{\nabla u}{|\nabla|}$  可转化为 (16) 式，奇异性消除，而原始 -对偶问题是为了合理地解释这种变换引入的。(这只是我的猜想，嘿嘿)

对于 (8) 式来说,

$$\int_{\Omega} |\nabla u| dx dy \quad (11)$$

$$= \int_{\Omega} \frac{\nabla u \nabla u}{|\nabla u|} dx dy \quad (12)$$

$$= \sup_{|\omega| \leq 1} \int_{\Omega} \nabla u \omega dx dy \quad (13)$$

$$= \sup_{|\omega| \leq 1} \left\{ \int_{\partial\Omega} \frac{\partial u}{\partial n} \omega dx dy - \int_{\Omega} u \operatorname{div} \omega dx dy \right\} \quad (14)$$

$$= \sup \left\{ - \int_{\Omega} u \nabla \omega dx dy : \omega \in (C_0^\infty(\Omega))^2, \|\omega\| \leq 1 \right\} \quad (15)$$

且  $\omega$  满足:

$$|\nabla u| \omega - \nabla u = 0 \quad (16)$$

(此处值得思考)。 $\omega$  把  $u$  映射为一个数, 故  $\omega$  为对偶向量。引入  $\omega$  后, 原始问题可表示为:

$$\min_u \sup_{\|\omega\| \leq 1} \Phi(u, \omega) \quad (17)$$

其中

$$\Phi(u, \omega) = \int_{\Omega} (-\alpha u \nabla \omega + \frac{1}{2} (Ku - z)^2) dx dy \quad (18)$$

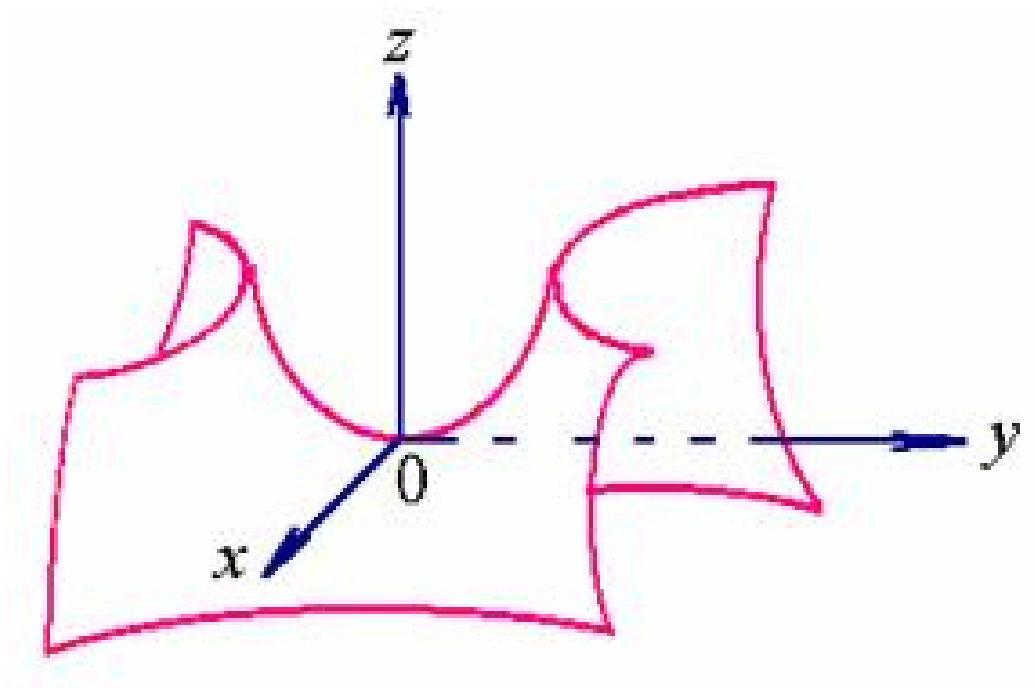
在凸分析中提到  $\Phi(u, \omega)$  有马鞍点, 故:

$$\min_u \sup_{\|\omega\| \leq 1} \Phi(u, \omega) = \sup_{\|\omega\| \leq 1} \min_u \Phi(u, \omega) \quad (19)$$

结合上图理解上述过程。

由此将原始问题转化为关于对偶向量  $\omega$  的对偶问题:

$$\sup_{\|\omega\| \leq 1} \Psi(\omega) \quad (20)$$



其中  $\Psi(\omega) = \min_u \Phi(u, \omega)$

## 6 牛顿法

### 6.1 对偶理论中的牛顿法

原始问题:  $\{\min c^T x; Ax = b, x \geq 0\}$  和对偶问题  $\{\max b^T \lambda; A^T \lambda + s = c, s \geq 0\}$  等价于下述问题:

$$\begin{aligned} A^T \lambda + s &= c \\ Ax &= b \\ x_i s_i &= 0 \\ (x, s) &\geq 0 \end{aligned} \tag{21}$$



用  $F(x, \lambda, s)$  来表示最优条件 (11):

$$\begin{bmatrix} A^T \lambda + s - c \\ Ax - b \\ x_i s_i \end{bmatrix} = 0 \quad (22)$$

在迭代求解时要确定前进方向。牛顿法在当前一步附近形成一个线性模型, 获得前进方向。它是通过解决下列线性方程来实现的:

$$J(x, \lambda, s) \begin{bmatrix} \Delta x \\ \Delta \lambda \\ \Delta s \end{bmatrix} = -F(x, \lambda, s) \quad (23)$$

其中  $J$  为  $F$  的雅可比矩阵。

$$J = \begin{bmatrix} 0 & A^T & I \\ A & 0 & 0 \\ S & 0 & X \end{bmatrix} \quad (24)$$

## 6.2 论文中的牛顿法

由第五章已知, 最原始的问题 (4) 等价解决关于  $u$  和  $\omega$  的原始 -对偶问题, 即解决欧拉 -拉格朗日方程:

$$\begin{aligned} -\alpha \nabla \omega + K^*(Ku - z) &= 0 \\ \omega \sqrt{|\nabla u|^2 + \beta} - \nabla u &= 0 \end{aligned} \quad (25)$$

记

$$F(u, \omega) = \begin{bmatrix} -\alpha \nabla \omega + K^*(Ku - z) \\ \omega \sqrt{|\nabla u|^2 + \beta} - \nabla u \end{bmatrix} \quad (26)$$

在推导  $F$  的雅可比矩阵前，先借助方向导数的概念来理解  $F$  的雅可比矩阵求法。函数  $f$  在点  $x$  处沿向量  $a$  的方向导数为：

$$D_a f(x) = \lim_{\tau \rightarrow 0} \frac{f(x + \tau a) - f(x)}{\tau} \quad (27)$$

如在 (14) 中的  $A^T$  可由方向导数推导而来，

$$D_\lambda J_1 = \lim_{\tau \rightarrow 0} \frac{A^T(\hat{\lambda} + \tau \lambda) + s - c - (A^T \hat{\lambda} + s - c)}{\tau} = A^T \lambda \quad (28)$$

对于论文中提到的雅可比矩阵， $u$  是关于  $x$ 、 $y$  的函数，当求  $u$  的偏导数时要涉及泛函导数的概念：对任何测试函数  $f$ ， $F$  关于  $\varphi$  的导数  $\frac{\delta F(\varphi)}{\delta \varphi}$  满足：

$$\left( \frac{\delta F(\varphi)}{\delta \varphi}, f \right) = \lim_{\tau \rightarrow 0} \frac{F(\varphi + \tau f) - F(\varphi)}{\tau} \quad (29)$$

下面分别推导  $F$  的雅可比矩阵的四项：记  $f_1 = -\alpha \nabla \omega + K^*(Ku - z)$ 、 $f_2 = \omega \sqrt{|\nabla u|^2 + \beta} - \nabla u$ ，用  $\hat{u}$ 、 $\hat{\omega}$  表示任意测试函数。

$$\begin{aligned} \left( \frac{\delta f_1(u, \omega)}{\delta \omega}, \hat{\omega} \right) &= \lim_{\tau \rightarrow 0} \frac{f_1(u, \omega + \tau \hat{\omega}) - f_1(u, \omega)}{\tau} \\ &= \lim_{\tau \rightarrow 0} \frac{[-\alpha \nabla (\omega + \tau \hat{\omega}) + K^*(Ku - z)] - [-\alpha \nabla \omega + K^*(Ku - z)]}{\tau} \\ &= -\alpha \nabla \hat{\omega} \end{aligned} \quad (30)$$

$$\begin{aligned} \left( \frac{\delta f_1(u, \omega)}{\delta u}, \hat{u} \right) &= \lim_{\tau \rightarrow 0} \frac{f_1(u + \tau \hat{u}, \omega) - f_1(u, \omega)}{\tau} \\ &= \lim_{\tau \rightarrow 0} \frac{[-\alpha \nabla \omega + K^*K(u + \tau \hat{u}) - K^*z] - [-\alpha \nabla \omega + K^*(Ku - z)]}{\tau} \\ &= K^*K \hat{u} \end{aligned} \quad (31)$$

$$\begin{aligned}
\left( \frac{\delta f_2(u, \omega)}{\delta \omega}, \hat{\omega} \right) &= \lim_{\tau \rightarrow 0} \frac{f_2(u, \omega + \tau \hat{\omega}) - f_2(u, \omega)}{\tau} \\
&= \lim_{\tau \rightarrow 0} \frac{[(\omega + \tau \hat{\omega}) | \nabla u | - \nabla u] - [\omega | \nabla u | - \nabla u]}{\tau} \\
&= | \nabla u | \hat{\omega}
\end{aligned} \tag{32}$$

$$\begin{aligned}
\left( \frac{\delta f_2(u, \omega)}{\delta u}, \hat{u} \right) &= \lim_{\tau \rightarrow 0} \frac{f_2(u + \tau \hat{u}, \omega)}{\tau} \\
&= \lim_{\tau \rightarrow 0} \frac{[\omega | \nabla (u + \tau \hat{u}) | - \nabla (u + \tau \hat{u})] - [\omega | \nabla u | - \nabla u]}{\tau}
\end{aligned} \tag{33}$$

$$\begin{aligned}
&\lim_{\tau \rightarrow 0} \frac{\nabla (u + \tau \hat{u}) - \nabla u}{\tau} = \nabla \hat{u} \\
&\lim_{\tau \rightarrow 0} \omega \frac{| \nabla (u + \tau \hat{u}) | - | \nabla u |}{\tau} \\
&= \lim_{\tau \rightarrow 0} \omega \frac{\sqrt{(\nabla u + \tau \hat{u})^T (\nabla u + \tau \hat{u})} - \sqrt{\nabla u^T \nabla u}}{\tau} \\
&= \lim_{\tau \rightarrow 0} \omega \frac{(\nabla u + \tau \nabla \hat{u})^T (\nabla u + \tau \nabla \hat{u}) - \nabla u^T \nabla u}{\tau [\sqrt{(\nabla u + \tau \nabla \hat{u})^T (\nabla u + \tau \nabla \hat{u})} + \sqrt{\nabla u^T \nabla u}]} \\
&= \lim_{\tau \rightarrow 0} \omega \frac{\tau \nabla \hat{u}^T \nabla u + \tau \nabla u^T \nabla \hat{u} + \tau^2 \nabla \hat{u}^T \nabla \hat{u}}{\tau [\sqrt{(\nabla u + \tau \nabla \hat{u})^T (\nabla u + \tau \nabla \hat{u})} + \sqrt{\nabla u^T \nabla u}]} \\
&= \frac{\omega \nabla u^T \nabla \hat{u}}{| \nabla u |}
\end{aligned} \tag{34}$$

将 (24) (25) 代入 (23) 中, 得:

$$\begin{aligned}
\left( \frac{\delta f_2(u, \omega)}{\delta u}, \hat{u} \right) &= \frac{\omega \nabla u^T \nabla \hat{u}}{| \nabla u |} - \nabla \hat{u} \\
&= \left( \frac{\omega \nabla u^T}{| \nabla u |} - I \right) \nabla \hat{u}
\end{aligned} \tag{35}$$

故  $F$  的雅可比矩阵为:

$$\begin{bmatrix} |\nabla u| & (\frac{\omega \nabla u^T}{|\nabla u|} - I) \nabla \\ -\alpha \nabla & K^* K \end{bmatrix} \quad (37)$$

同理可推出论文中  $u$  系统的线性化形式 (3.3)。

## 参考文献

- [1] <http://zh.wikipedia.org/wiki/>
- [2] 梁灿彬, 周彬. 《微分几何入门及广义相对论》. 科学出版社, 2000
- [3] Krishnan D, Lin P, Yip A. A primal-dual active-set method for non-negativity constrained total variation deblurring problems. IEEE Trans Image Process 16(11):2766-2777, 2007
- [4] Xiaoqun Zhang, Martin Burger, Stanley Osher, A Unified Primal-Dual Algorithm Framework Based on Bregman Iteration. Journal of Scientific Computing, Volume 46, Number 1, 20-46, 2011
- [5] Peijun Chen, Jianguo Huang and Xiaoqun Zhang, A Primal-Dual Fixed Point Algorithm for Convex Separable Minimization with Applications to Image Restoration. Inverse problems, 29 (2), 2013
- [6] Ernie Esser, Xiaoqun Zhang, Tony F. Chan, A General Framework for a Class of First Order Primal-Dual Algorithms for Convex Optimization in Imaging Science. SIAM J. Imaging Sci., 3(4), 1015-1046, 2010
- [7] Yiqiu Dong, Michael Hintermuller, Marrick Neri, An Efficient Primal-Dual Method for L1TV Image Restoration. SIAM J. Imaging SCIENCES, Vol 2, No. 4, pp.1168-1189.

- [8] You-Wei Wen, Raymond H. Chan, Andy M. Yip, A Primal-Dual Method for Total Variation-Based Wavelet Domain Inpainting, IEEE Trans Image Process. 2012 Jan;21(1):106-14.
- [9] N. Megiddo, Pathways to the optimal set in linear programming, in: N. Megiddo(Eds.), Progress in Mathematical Programming:Interior-Point and Related Methods, Springer, New York, 1989, 131-158.
- [10] S. Mehrotra, On the implementation of a primal-dual interior point method, SIAM Journal on Optimization, 1992, 2:575-601.
- [11] Y. Nesterov, A. Nemirovski, Interior point polynomial methods in convex programming: theory and application, SIAM, Philadelphia, PA, 1994.
- [12] Y. Nesterov, M. Todd, Self-scaled barriers and interior-point methods for convex programming, Mathematical Methods of Operations Research, 1997, 22: 1-42.