

非凸优化中一种带非单调线搜索的 惯性邻近算法

刘海玉

河北工业大学, 理学院, 天津
Email: 1023380197@qq.com

收稿日期: 2021年2月23日; 录用日期: 2021年3月19日; 发布日期: 2021年3月26日

摘 要

本文考虑一类目标函数由可微(可能非凸)函数和凸(可能非光滑)函数组成的极小化问题。惯性邻近(iPiano)算法是解决这类问题的一种有效而重要的方法。通过引入非单调线搜索, 提出了非单调线搜索的iPiano (iPiano-nml)算法。通过证明说明了由iPiano-nml生成的序列的任何聚点都是一个稳定点。最后, 对图像处理问题进行了数值实验来说明新算法的理论结果。

关键词

非单调邻近梯度法, 非凸, 非光滑, 图像降噪

An Inertial Proximal Algorithm with Non-Monotone Line Search for Non-Convex Optimization

Haiyu Liu

School of Science, Hebei University of Technology, Tianjin
Email: 1023380197@qq.com

Received: Feb. 23rd, 2021; accepted: Mar. 19th, 2021; published: Mar. 26th, 2021

Abstract

We consider a class of minimization problems whose objective function is composed of a differentiable (possibly nonconvex) function and a convex (possibly nonsmooth) function. The iner-

tial proximal algorithm is a common and important method for this kind of problems. By incorporating the non-monotone line search, iPiano algorithm with non-monotone line search is proposed. We show that any cluster point of sequence which is generated by iPiano-nml is a stationary point. Finally, we perform numerical experiments on the image processing problems to illustrate our theoretical results.

Keywords

Non-Monotone Proximal Gradient Method, Nonconvex, Non-Smooth, Image Denoising

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

本文考虑下列最小化问题:

$$\min_{x \in R^n} F(x) = f(x) + g(x), \tag{1}$$

其中, $f(x): R^N \rightarrow R$ 是一个光滑(可能非凸)且为梯度李普希兹连续函数, $g(x): R^N \rightarrow R$ 是一个正常下半连续凸(可能非光滑)函数。另外 $F(x)$ 是一个强制函数且有下界。

问题(1)出现在很多领域, 如机器学习[1], 信号处理[2]等。解决上述问题的经典方法是邻近梯度法[3]:

$$x^{k+1} \in \arg \min_x \left\{ \langle \nabla f(x^k), x \rangle + \frac{\alpha}{2} \|x - x^k\|^2 + g(x) \right\},$$

其中 α 是步长。然而, 邻近梯度算法在解决一些问题时, 它的收敛速度较慢, 于是人们逐渐地寻找一些策略来改善这样的问题。常见的一个有效的策略是加速方法, 因此在上述邻近梯度法中, 通过引入外推项, 一些文献提出了加速邻近梯度算法[4], FISTA [5]算法等。加速邻近梯度算法如下:

$$\begin{aligned} x^{k+1} &\in \arg \min_x \left\{ \langle \nabla f(y^k), x \rangle + \frac{\alpha}{2} \|x - y^k\|^2 + g(x) \right\}, \\ y^k &= x^k + \beta_k (x^k - x^{k-1}), \end{aligned}$$

其中 α 是步长, β_k 是外推参数。这些算法的收敛速度为 $O\left(\frac{1}{k^2}\right)$ 。除此自外, 惯性项也是一类能加速算法的有效策略, 通过将邻近梯度法和惯性项结合得到下列惯性邻近算法(iPiano) [6]:

$$x^{k+1} = (I + \alpha \partial g)^{-1} (x^k - \alpha \nabla f(x^k) + \beta (x^k - x^{k-1})), \tag{2}$$

其中, α 是步长, β 是惯性因子。iPiano 在实际运行中收敛速度比邻近梯度法快, 广泛地应用于非凸问题[7] [8]。另一方面, 线搜索也是一项能加速原始算法的有效策略, 其非单调线搜索有更好的数值表现。非单调邻近梯度法(NPG) [9]采用了下列线搜索:

$$u \in \arg \min_x \left\{ \langle \nabla f(x^k), x - x^k \rangle + \frac{L}{2} \|x - x^k\|^2 + g(x) \right\},$$

$$F(u) \leq \max_{[k-M]_+ \leq i \leq k} F(x^i) - \frac{c}{2} \|u - x^k\|^2,$$

其中 c 和 M 都大于 0。该线搜索在每次迭代中通过选取 k 与 $k-M$ 之间的最大值来满足条件，从而保证目标函数有更大的下降。

本文在惯性邻近算法的基础上，将其与 NPG 结合，提出一种带非单调线搜索的惯性邻近算法，并将其运用于图像处理，通过信噪比来说明新算法的有效性。

2. 预备定义

本节给出一些基本的定义。本文考虑欧式空间且维数 $N \geq 1$ 。

定义 1: 令 $g(x)$ 为一个正常下半连续凸函数，那么邻近算子定义为：

$$(I + \alpha \partial g)^{-1}(\hat{x}) = \arg \min_x \frac{\|x - \hat{x}\|^2}{2} + \alpha g(x).$$

引理 1: $f(x): R^N \rightarrow R$ 为一个光滑函数且为 L 梯度李普希兹连续。则对于 $\forall x, y \in R^N$ ， $\exists L > 0$ ，下式成立：

$$\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\|.$$

引理 2: $f(x): R^N \rightarrow R$ 为一个光滑函数且为 L 梯度李普希兹连续。则对于 $\forall x, y \in R^N$ ，我们成立以下下降引理

$$f(x) \leq f(y) + \langle \nabla f(y), x - y \rangle + \frac{L}{2} \|x - y\|^2.$$

命题 1: 令 F, f, g 满足上述定义，则对于 $x \in \text{dom}F$ ，我们有下式成立：

$$\partial F(x) = \nabla f(x) + \partial g(x).$$

3. 算法和收敛性

本节给出 iPiano-nml 算法和收敛性分析。首先我们引入一个辅助序列：

$$H_\delta(x, y) = f(x) + g(x) + \delta \|x - y\|^2.$$

易知当时 $x = y$ ，我们有 $H_\delta(x, y) = f(x) + g(x)$ 。

3.1. iPiano-nml 算法

iPiano-nml 算法：

步骤 0: 选择 $x^0 \in \text{dom}F, x^0 = x^{-1}$ 。 c_1, c_2 接近 0。 $L_{\max} \geq L_{\min} > 0, \tau > 1, c > 0, M \geq 0$ 。

步骤 k: 1) 选择 $L_k \in [L_{\min}, L_{\max}]$, $y^k = x^k + \beta_k(x^k - x^{k-1})$ 。

1a) 求解子问题：

$$u \in \arg \min_x \left\{ g(x) + \langle \nabla f(x^k), x \rangle + \frac{1}{2\alpha_k} \|x - y^k\|^2 \right\}.$$

1b) 如果满足

$$H_{\delta_{k+1}}(u, x^k) \leq \max_{[k-M]_+ \leq i \leq k} H_{\delta_i}(x^i, x^{i-1}) - \frac{c}{2} \|u - x^k\|^2,$$

那么进行步骤 2)。

1c) 令 $L_k \leftarrow \tau L_k$, 继续步骤 1a)。

2) 令 $x^{k+1} \leftarrow u, k \leftarrow k+1$ 然后进行步骤 1)。

其中序列满足 $\alpha_k \geq c_1, \beta_k \geq 0, \delta_k \geq \gamma_k \geq c_2$ 。 δ_k 单调递减, 另有:

$$\delta_k = \frac{1}{\alpha_k} - \frac{L_k}{2} - \frac{\beta_k}{2\alpha_k}, \gamma_k = \frac{1}{\alpha_k} - \frac{L_k}{2} - \frac{\beta_k}{\alpha_k}.$$

3.2. 收敛性分析

引理 1: 对于 $k \geq 0$, 存在 $\delta_k \geq \gamma_k, \beta_k \in [0,1), \alpha_k < \frac{2(1-\beta_k)}{L_k}$ 。另外给定 $L_k \geq 0$, 存在 α_k, β_k 满足 δ_k 单调递减。

证明: 该引理证明参考文献[6]。

引理 2: 序列 $(H_{\delta_k}(x^k, x^{k-1}))_{k=0}^\infty$ 单调递减且收敛, 且我们有

$$H_{\delta_{k+1}}(x^{k+1}, x^k) \leq H_{\delta_k}(x^k, x^{k-1}) - \gamma_k \|x^k - x^{k-1}\|^2.$$

证明: 该引理证明参考文献[6]。

定理 1:

- 1) 序列 $\{x^k\}$ 有界。
- 2) $\lim_{k \rightarrow \infty} \|x^k - x^{k-1}\| = 0$ 。
- 3) 任意序列的聚点都是稳定点。

证明:

1) 由引理 2 可知整个序列 $\{x^k\}$ 都包含于下列水平集合:

$$\{x \in R^N : \underline{F} \leq F(x) \leq H_{\delta_0}(x^0, x^{-1})\}.$$

那么由 $\underline{F} = \inf F(x) > -\infty$, 我们可推知序列 $\{x^k\}$ 有界。

2) 显然, 我们有:

$$\begin{aligned} & H_{\delta_{k+1}}(x^{k+1}, x^k) - \max_{[k-M]_+ \leq i \leq k} H_{\delta_i}(x^i, x^{i-1}) \\ &= f(x^{k+1}) + g(x^{k+1}) + \delta_{k+1} \|x^{k+1} - x^k\|^2 - \max_{[k-M]_+ \leq i \leq k} \{f(x^i) + g(x^i) + \delta_i \|x^i - x^{i-1}\|^2\} \\ &\leq f(x^{k+1}) + g(x^{k+1}) + \delta_{k+1} \|x^{k+1} - x^k\|^2 - f(x^k) - g(x^k) - \delta_k \|x^k - x^{k-1}\|^2 \\ &= H_{\delta_{k+1}}(x^{k+1}, x^k) - H_{\delta_k}(x^k, x^{k-1}). \end{aligned}$$

由引理 2 可知:

$$H_{\delta_{k+1}}(x^{k+1}, x^k) - \max_{[k-M]_+ \leq i \leq k} H_{\delta_i}(x^i, x^{i-1}) \leq -\gamma_k \|x^k - x^{k-1}\|^2.$$

将上式两边从 $k = 0$ 加到 N 。则由引理 2 和 M 的定义我们有

$$\begin{aligned} 0 &< \sum_{k=0}^N \gamma_k \|x^k - x^{k-1}\|^2 \leq \sum_{k=0}^N \max_{[k-M]_+ \leq i \leq k} H_{\delta_i}(x^i, x^{i-1}) - H_{\delta_{k+1}}(x^{k+1}, x^k) \\ &= (M-1)H_{\delta_0}(x^0, x^{-1}) - \dots - H_{\delta_{N-M+2}}(x^{N-M+2}, x^{N-M+1}) \\ &\quad - H_{\delta_{N-M+3}}(x^{N-M+3}, x^{N-M+2}) - \dots - H_{\delta_N}(x^N, x^{N-1}) - H_{\delta_{N+1}}(x^{N+1}, x^N) < \infty, \end{aligned}$$

从而我们可以推断出 $\lim_{k \rightarrow \infty} \|x^k - x^{k-1}\| = 0$ 。

3) 令 x^* 为序列的聚点且存在 $\{x^{k_j}\}$ 使得 $\lim_{j \rightarrow \infty} x^{k_j} = x^*$ 。那么由一阶最优条件，我们有

$$-\frac{1}{\alpha_{k_j}} \|x^{k_j+1} - y^{k_j}\| \in \partial g(x^{k_j+1}) + \nabla f(x^{k_j}).$$

由 $y^{k_j} = x^{k_j} + \beta_{k_j}(x^{k_j} - x^{k_j-1})$ ，我们可以得到

$$-\frac{1}{\alpha_{k_j}} \|x^{k_j+1} - x^{k_j} - \beta_{k_j}(x^{k_j} - x^{k_j-1})\| \in \partial g(x^{k_j+1}) + \nabla f(x^{k_j}).$$

由函数 $g(x)$ 的凸性， $\nabla f(x)$ 连续性以及 $\|x^k - x^{k-1}\| = 0$ ，我们有

$$0 \in \nabla f(x^*) + \partial g(x^*),$$

这就说明了 x^* 是算法的稳定点。

4. 数值实验

本节将算法用于数值实验，运行环境为 MATLAB R2014a 2.80 GHz CPUs。我们先简单介绍一下问题模型。

问题模型为马尔科夫随机场：

$$\min_{u \in R^N} \sum_{i=1}^{N_f} \zeta_i \phi(K_i u) + g(u, u^0),$$

其中 u 是所求图像信息， u^0 是噪声图像信息， ϕ 是非凸罚函数，表达式为：

$$\phi(K_i u) = \sum_p \varphi((K_i u)_p),$$

其中 K_i 是学习率，是基于 ζ_i 权重的线性算子， N_f 是滤子数，本文考虑 $7 * 7$ 的滤波器[6]，且 $K_i u = k_i * u$ 。

对于 $\varphi(t)$ ，我们考虑学生 t 分布：

$$\varphi(t) = \log(1 + t^2).$$

我们考虑高斯噪声，其中函数模型为 $g(u, u^0) = \frac{\lambda}{2} \|u - u^0\|^2$ 。

对于终止条件我们选择下述能量差：

$$\varepsilon_k = F_k - F_*,$$

其中， F_* 是真实值(数值由[6]给出)， F_k 是每次迭代的值，两者做差即为能量差 ε_k 。

为了增强对比效果，我们用 iPiano-nml 和 iPiano-Backtracking 作比较。iPiano-Backtracking 为带回溯线搜索的 iPiano 算法：

$$f(x^{k+1}) \leq f(x^k) + \langle \nabla f(x^k), x^{k+1} - x^k \rangle + \frac{L_k}{2} \|x^{k+1} - x^k\|^2 \quad (k \in N).$$

当满足回溯线搜索时，令 $L_{k+1} = L_k / 1.05$ 否则令 $L_{k+1} = L_k * 1.2$ 。

在 iPiano-nml 中，令 L_k 为 BB 步长[10]：

$$L_k = \begin{cases} \min \left(\max \left(\frac{\|s_{k-1}\|^2}{s_{k-1}^T l_{k-1}}, L_{\min} \right), L_{\max} \right) & (s_{k-1}^T l_{k-1} \neq 0) \\ L_{\max}, & \text{otherwise.} \end{cases}$$

另外参数选取为:

$$L_1 = 3, \alpha_k = \frac{2 * (1 - \beta)}{L_k} * 0.99, \beta = 0.8, \lambda = 1, c = 4.8, \tau = 1.6, M = 2.$$

在 iPiano-Backtracking 中, 令参数选取为:

$$L_1 = 3, \alpha_k = \frac{2 * (1 - \beta)}{L_k} * 0.99, \beta = 0.8, \lambda = 1.$$

接下来我们给出峰值信噪比与迭代变化关系图:

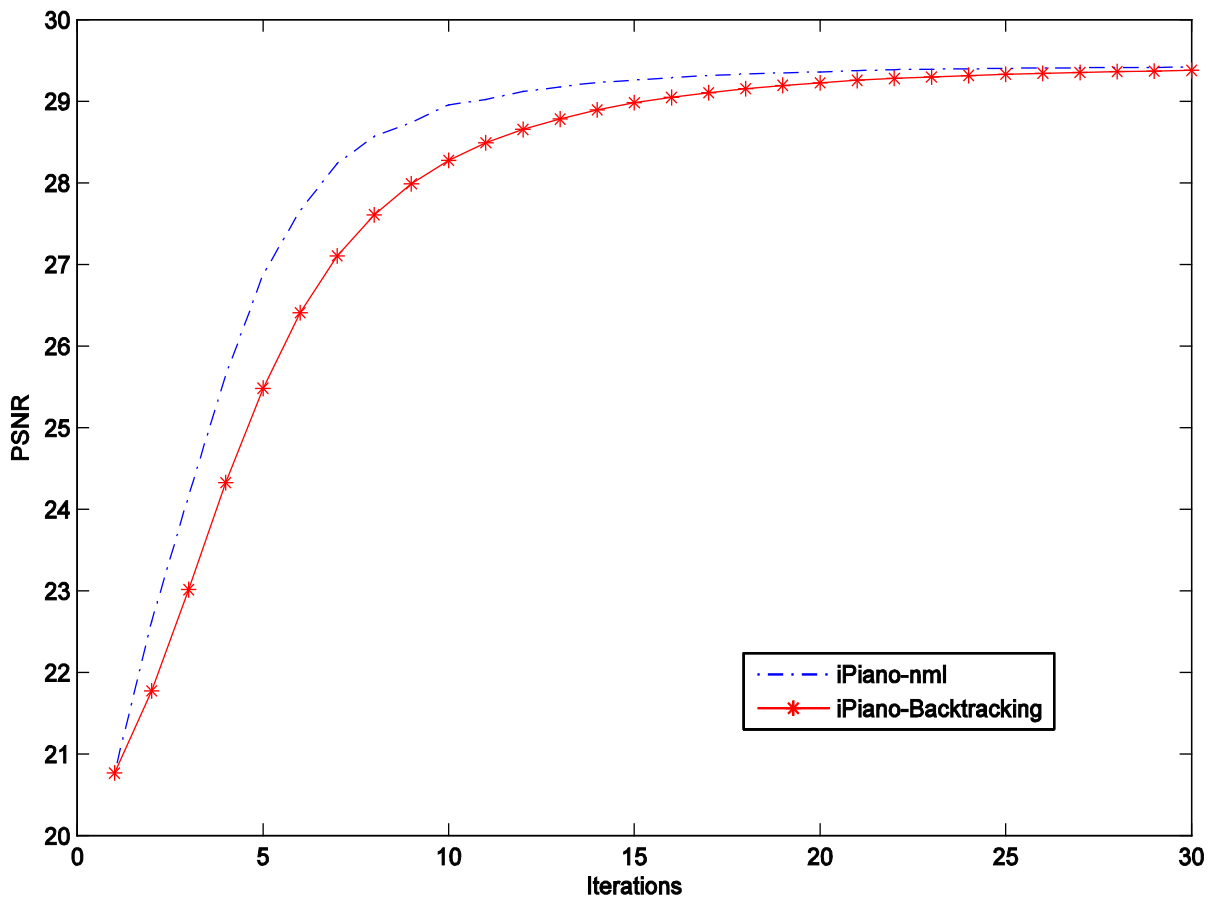


Figure 1. Comparison of two algorithms

图 1. 两个算法效果对比图

由图 1 中可知在过程中 iPiano-nml 比 iPiano-Backtracking 效果表现好。

下面给出效果对比图:

图 2 出了效果对比。上方左图为原始图像, 右图为噪声图像。下方左图为 iPiano-nml 去噪效果, 右图为 iPiano-Backtracking 去噪效果。图 2 验证了 iPiano-nml 算法的有效性。

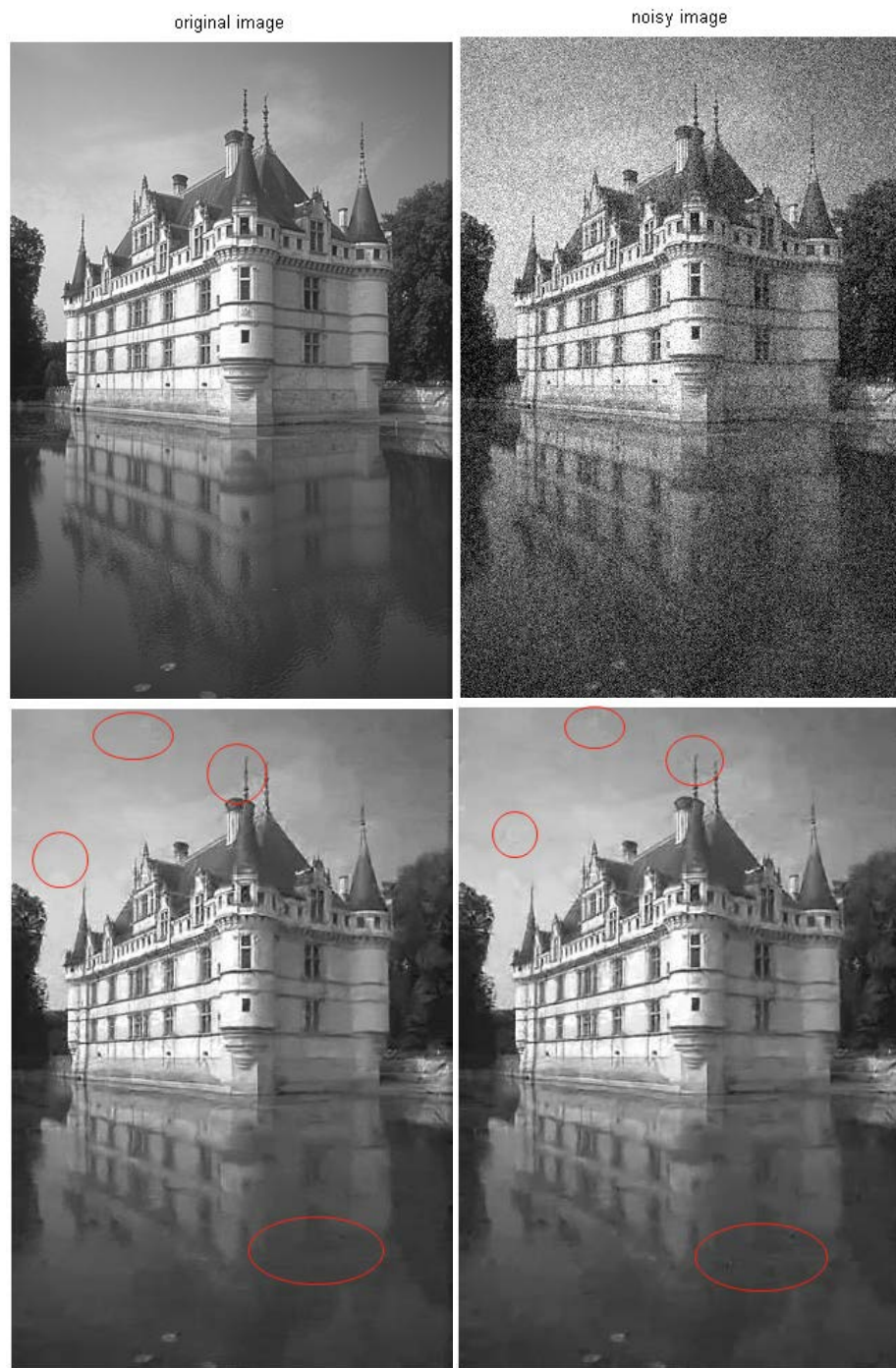


Figure 2. Effect comparison
图 2. 效果对比

5. 结论

本文在考虑一类非凸非光滑问题中借鉴了 NPG 的思想，将非单调线搜索整合到 iPiano 中，提出了带非单调技术的 iPiano 算法，并通过一定的条件证明了新算法的收敛性。本文将新算法运用于图像降噪实验，通过与原算法的对比说明新算法的有效性。以后我们的工作将研究新算法的收敛速度。

参考文献

- [1] Curtis, F.E. and Scheinberg, K. (2017) Optimization Methods for Supervised Machine Learning: From Linear Models to Deep Learning. In: *Leading Developments from INFORMS Communities*, 89-114. <https://doi.org/10.1287/educ.2017.0168>
- [2] Combettes, P.L. and Pesquet, J.C. (2011) Proximal Splitting Methods in Signal Processing. In: Bauschke, H., Burachik, R., Combettes, P., Elser, V., Luke, D. and Wolkowicz, H. (Eds.), *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, Springer, New York, NY, 185-212. https://doi.org/10.1007/978-1-4419-9569-8_10
- [3] Boş, R.I., Csetnek, E.R. and Hendrich, C. (2015) Inertial Douglas-Rachford Splitting for Monotone Inclusion Problems. *Applied Mathematics and Computation*, **256**, 472-487. <https://doi.org/10.1016/j.amc.2015.01.017>
- [4] Wen, B. and Xue, X. (2019) On the Convergence of the Iterates of Proximal Gradient Algorithm with Extrapolation for Convex Nonsmooth Minimization Problems. *Journal of Global Optimization*, **75**, 767-787. <https://doi.org/10.1007/s10898-019-00789-8>
- [5] Beck, A. and Teboulle, M. (2009) A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM journal on Imaging Sciences*, **2**, 183-202. <https://doi.org/10.1137/080716542>
- [6] Ochs, P., Chen, Y., Brox, T., *et al.* (2014) iPiano: Inertial Proximal Algorithm for Nonconvex Optimization. *SIAM Journal on Imaging Sciences*, **7**, 1388-1419. <https://doi.org/10.1137/130942954>
- [7] Ochs, P. (2019) Unifying Abstract Inexact Convergence Theorems and Block Coordinate Variable Metric iPiano. *SIAM Journal on Optimization*, **29**, 541-570. <https://doi.org/10.1137/17M1124085>
- [8] Ochs, P. (2018) Local Convergence of the Heavy-Ball Method and iPiano for Non-Convex Optimization. *Journal of Optimization Theory and Applications*, **177**, 153-180. <https://doi.org/10.1007/s10957-018-1272-y>
- [9] Chen, X., Lu, Z. and Pong, T.K. (2016) Penalty Methods for a Class of Non-Lipschitz Optimization Problems. *SIAM Journal on Optimization*, **26**, 1465-1492. <https://doi.org/10.1137/15M1028054>
- [10] Barzilai, J. and Borwein, J.M. (1988) Two-Point Step Size Gradient Methods. *IMA Journal of Numerical Analysis*, **8**, 141-148. <https://doi.org/10.1093/imanum/8.1.141>