Proximal Algorithms in Signal Processing

Mostafa Sadeghi

Electrical Engineering Department
Sharif University of Technology
Tehran, IRAN

March 2017

Overview

1. Proximal Algorithms
   - Proximal Point Method
   - Forward-backward Splitting
   - Backward-backward Splitting
   - Applications

Proximal Algorithms (convex)

Proximal Algorithms

Authors	Neal Parikh, Stephen P Boyd
Publication date	2014/1/13
Journal	Foundations and Trends in optimization
Volume	1
Issue	3
Pages	127-239
Description	Abstract This monograph is about a class of optimization algorithms called proximal algorithms. Much like Newton's method is a standard tool for solving unconstrained smooth optimization problems of modest size, proximal algorithms can be viewed as an analogous tool for nonsmooth, constrained, large-scale, or distributed versions of these problems. They are very generally applicable, but are especially well-suited to problems of substantial recent interest involving large or high-dimensional datasets. Proximal methods sit at a ...
Total citations	Cited by 819

Proximal Algorithms (non-convex)

Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods

Authors	Hedy Attouch, Jérôme Bolte, Benar Fux Svaiter
Publication date	2013/2/1
Journal	Mathematical Programming Volume 137 Issue 1-2 Pages 91-129 Publisher Springer-Verlag
Total citations	Cited by 267

Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality

Authors	Hédy Attouch, Jérôme Bolte, Patrick Redont, Antoine Soubeyran
Publication date	2010/5
Journal	Mathematics of Operations Research Volume 35 Issue 2 Pages 438-457 Publisher INFORMS
Total citations	Cited by 269

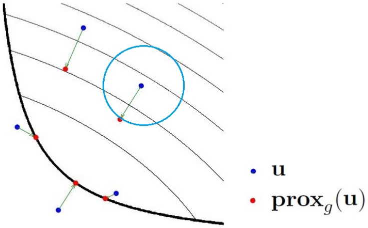
Proximal Mapping

Proximal mapping

$g: \ \mathbf{dom}_g \to \mathbb{R} \cup \{+\infty\}$: proper, lower-semicontinuous

$$\boxed{\mathbf{prox}_g(\mathbf{u}) = \underset{\mathbf{x} \in \mathbf{dom}_g}{\operatorname{argmin}} \ g(\mathbf{x}) + \frac{1}{2}\|\mathbf{x} - \mathbf{u}\|_2^2}$$

$$\underset{\mathbf{x} \in \mathbf{dom}_g}{\operatorname{argmin}} \ g(\mathbf{x}) \quad \text{s.t.} \quad \|\mathbf{x} - \mathbf{u}\|_2 \le \tau$$



Proximal Mapping

Interpretations:

- Generalized projection: $g(\mathbf{x}) = \delta_{\mathcal{C}}(\mathbf{x}) \triangleq \begin{cases} 0 & \mathbf{x} \in \mathcal{C} \\ \infty & \mathbf{x} \notin \mathcal{C} \end{cases}$

$$\boxed{\mathbf{prox}_g(\mathbf{u}) = \mathcal{P}_{\mathcal{C}}(\mathbf{u})}$$

- Gradient step: g is smooth and $\lambda > 0$ is small

$$\boxed{\begin{aligned} \mathbf{prox}_{\lambda g}(\mathbf{u}) &= \underset{\mathbf{x} \in \mathbf{dom}_g}{\operatorname{argmin}} \ \lambda g(\mathbf{x}) + \frac{1}{2}\|\mathbf{x} - \mathbf{u}\|_2^2 \\ &\simeq \mathbf{u} - \lambda \nabla g(\mathbf{u}) \end{aligned}}$$

Proximal Mapping

Properties:

1. Postcomposition: If $f(\mathbf{x}) = \alpha\phi(\mathbf{x}) + \beta$ then $\mathbf{prox}_f(\mathbf{x}) = \mathbf{prox}_{\alpha\phi}(\mathbf{x})$
2. Precomposition: If $f(\mathbf{x}) = \phi(\alpha\mathbf{x} + \beta)$ then $\mathbf{prox}_f(\mathbf{x}) = 1/\alpha\big(\mathbf{prox}_{\alpha^2\phi}(\alpha\mathbf{x} + \beta) - \beta\big)$
3. Moreau decomposition: $\mathbf{x} = \mathbf{prox}_f(\mathbf{x}) + \mathbf{prox}_{f^*}(\mathbf{x})$ $\quad f^*(\mathbf{y}) = \sup_{\mathbf{x}} \ (\mathbf{y}^T\mathbf{x} - f(\mathbf{x}))$

Example: $f(\mathbf{x}) = \|\mathbf{x}\|_\infty$

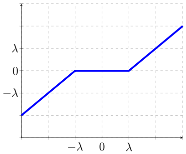
$$\mathbf{prox}_f(\mathbf{x}) = \mathbf{x} - \mathbf{prox}_{f^*}(\mathbf{x})$$

$$f^*(\mathbf{x}) = \delta_{\mathcal{C}}(\mathbf{x}), \quad \mathcal{C} = \{\mathbf{x} \in \mathbb{R}^n \mid \ \|\mathbf{x}\|_1 \leq 1\}$$

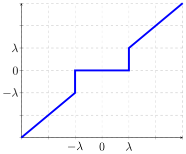
$$\boxed{\mathbf{prox}_{\|.\|_\infty}(\mathbf{x}) = \mathbf{x} - \mathbf{prox}_{\delta_{\mathcal{C}}}(\mathbf{x})}$$

Examples:

- Soft-thresholding: $g(\mathbf{x}) = \lambda \|\mathbf{x}\|_1$



- Hard-thresholding: $g(\mathbf{x}) = \lambda \|\mathbf{x}\|_0$



- Mixed norms: $\ell_{2,1}$, $\ell_{2,0}, \cdots$

Proximal Point Method

$$\min_{\mathbf{x}\in\mathbb{R}^n} \; f(\mathbf{x})$$

$$\boxed{\mathbf{x}_{k+1} = \operatorname*{argmin}_{\mathbf{x}\in\mathbb{R}^n} \; f(\mathbf{x}) + \frac{1}{2\alpha_k}\|\mathbf{x}-\mathbf{x}_k\|_2^2}$$

- smooth f: $\quad \mathbf{x}_{k+1} \simeq \mathbf{x}_k - \alpha_k \nabla f(\mathbf{x}_k)$
- non-smooth f: $\quad \mathbf{x}_{k+1} = \mathbf{prox}_{\alpha_k f}(\mathbf{x}_k)$

Forward-backward Splitting

$$\min_{\mathbf{x}\in\mathbb{R}^n} \; f(\mathbf{x}) + g(\mathbf{x})$$

- $f: \; \mathbf{dom}_f \to \mathbb{R}$ smooth (convex/non-convex)
- $g: \; \mathbf{dom}_g \to \mathbb{R} \cup \{+\infty\}$ non-smooth (convex/non-convex)

Examples:

- Compressed sensing: $\min_{\mathbf{x}} \; \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_0$
- Low-rank recovery: $\min_{\mathbf{X}} \; \frac{1}{2}\|\mathbf{Y} - \mathcal{A}(\mathbf{X})\|_F^2 + \lambda\|\mathbf{X}\|_*$
- Dictionary learning: $\min_{\mathbf{X},\mathbf{D}} \; \frac{1}{2}\|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 + \delta_x(\mathbf{X}) + \delta_d(\mathbf{D})$

A Key Lemma

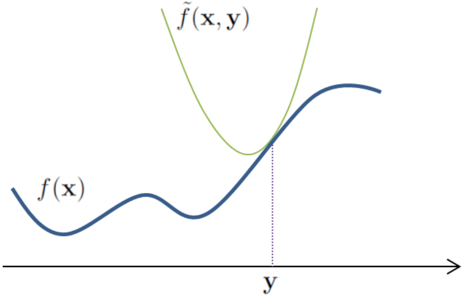
Descent lemma

$f: \mathbf{dom}_f \to \mathbb{R}$, smooth and L-gradient Lipschitz*, $\mu \in (0, 1/L]$

$$\forall \mathbf{x}, \mathbf{y} \in \mathrm{dom} f : \quad f(\mathbf{x}) \leq \underbrace{f(\mathbf{y}) + \nabla f(\mathbf{y})^T(\mathbf{x} - \mathbf{y}) + \frac{1}{2\mu}\|\mathbf{x} - \mathbf{y}\|_2^2}_{\tilde{f}(\mathbf{x},\mathbf{y})}$$

*$\forall x, y \in \mathrm{dom} f$:

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2 \leq L\|\mathbf{x} - \mathbf{y}\|_2$$



Forward-backward Splitting

$$\min_{\mathbf{x}\in\mathbb{R}^n} \; f(\mathbf{x}) + g(\mathbf{x})$$

$$\boxed{\mathbf{x}_{k+1} = \underset{\mathbf{x}}{\operatorname{argmin}} \; \tilde{f}(\mathbf{x}, \mathbf{x}_k) + g(\mathbf{x})}$$

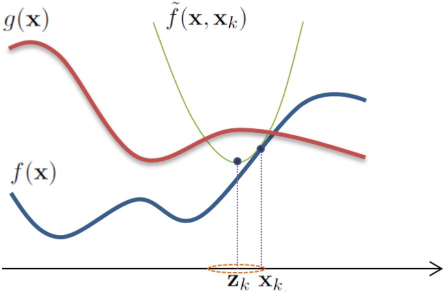
$$\mathbf{x}_{k+1} = \underset{\mathbf{x}}{\operatorname{argmin}} \; \frac{1}{2}\|\mathbf{x} - \big(\mathbf{x}_k - \mu\nabla f(\mathbf{x}_k)\big)\|_2^2 + \mu \cdot g(\mathbf{x})$$

1. Forward step:

$$\mathbf{z}_k = \mathbf{x}_k - \mu\nabla f(\mathbf{x}_k)$$

2. Backward step:

$$\mathbf{x}_{k+1} = \mathbf{prox}_{\mu\cdot g}(\mathbf{z}_k)$$



Forward-backward Splitting

Example 1: Iterative Shrinkage-Thresholding

$$\min_{\mathbf{x}\in\mathbb{R}^n} \underbrace{\frac{1}{2}\|\mathbf{y}-\mathbf{D}\mathbf{x}\|_2^2}_{f(\mathbf{x})} + \underbrace{\lambda\|\mathbf{x}\|_1}_{g(\mathbf{x})} \rightarrow \boxed{\mathbf{x}_{k+1} = \mathbf{prox}_{\mu\lambda\|.\|_1}\left(\mathbf{x}_k - \mu\nabla f(\mathbf{x}_k)\right)}$$

Example 2: Smoothed ℓ_0 (SL0)

$$\min_{\mathbf{x}\in\mathbb{R}^n} \sum_{i=1}^{n}\left(1-\exp(-\frac{x_i^2}{\sigma^2})\right) \quad \text{s.t.} \quad \|\mathbf{y}-\mathbf{D}\mathbf{x}\|_2 \leq \epsilon$$

$$\min_{\mathbf{x}\in\mathbb{R}^n} \underbrace{\sum_{i=1}^{n}\left(1-\exp(-\frac{x_i^2}{\sigma^2})\right)}_{f(\mathbf{x})} + \underbrace{\delta_{\mathcal{C}}(\mathbf{x})}_{g(\mathbf{x})} \rightarrow \boxed{\mathbf{x}_{k+1} = \mathbf{prox}_{\delta_{\mathcal{C}}}\left(\mathbf{x}_k - \mu\nabla f(\mathbf{x}_k)\right)}$$

Backward-backward Splitting

$$\min_{\mathbf{x}\in\mathbb{R}^n} \; g(\mathbf{x}) + h(\mathbf{x})$$

- $g: \; \mathbf{dom}_g \to \mathbb{R} \cup \{+\infty\}$ non-smooth (convex/non-convex)
- $h: \; \mathbf{dom}_h \to \mathbb{R} \cup \{+\infty\}$ non-smooth (convex/non-convex)

$$\boxed{\mathbf{x}_{k+1} = \mathbf{prox}_g\left(\mathbf{prox}_h(\mathbf{x}_k)\right)}$$

Example. Find $\mathbf{x} \in \mathcal{C}_1 \cap \mathcal{C}_2$

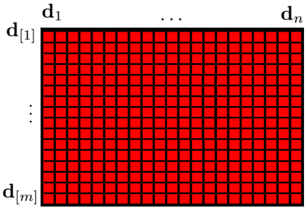
$$\boxed{\min_{\mathbf{x}\in\mathbb{R}^n} \; \delta_{\mathcal{C}_1}(\mathbf{x}) + \delta_{\mathcal{C}_2}(\mathbf{x})}$$

$$\delta_{\mathcal{C}}(\mathbf{x}) \triangleq \begin{cases} 0 & \mathbf{x} \in \mathcal{C} \\ \infty & \mathbf{x} \notin \mathcal{C} \end{cases}$$

Backward-backward Splitting

Example 1: Designing normalized-column, orthogonal-row matrices (unit-norm tight frames)

- $\mathcal{C}_1 = \{\mathbf{D} \in \mathbb{R}^{m \times n} \mid \forall i : \; \|\mathbf{d}_i\|_2 = 1\}$
- $\mathcal{C}_2 = \left\{\mathbf{D} \in \mathbb{R}^{m \times n} \mid \forall i \neq j : \; \mathbf{d}_{[i]}^T \mathbf{d}_{[j]} = 0\right\}$



$$\mathbf{D}_{k+1} = \underbrace{\mathbf{prox}_{\delta_{\mathcal{C}_2}}}_{\text{SVD}} \left(\underbrace{\mathbf{prox}_{\delta_{\mathcal{C}_1}}}_{\text{normalization}} (\mathbf{D}_k) \right)$$

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \to \mathbf{prox}_{\delta_{\mathcal{C}_2}}(\mathbf{A}) = \mathbf{U}\mathbf{V}^T$$

Example 2: Sparse recovery by ℓ_0 minimization

$$\boxed{\min_{\mathbf{x}\in\mathbb{R}^n} \ \|\mathbf{x}\|_0 \quad \text{s.t.} \quad \|\mathbf{y}-\mathbf{D}\mathbf{x}\|_2 \le \epsilon}$$

$$\min_{\mathbf{x}\in\mathbb{R}^n} \ \underbrace{\|\mathbf{x}\|_0}_{g(\mathbf{x})} + \underbrace{\delta_{\mathcal{C}}(\mathbf{x})}_{h(\mathbf{x})} \qquad \mathcal{C} = \{\mathbf{x}\in\mathbb{R}^n \mid \ \|\mathbf{y}-\mathbf{D}\mathbf{x}\|_2 \le \epsilon\}$$

$$\boxed{\mathbf{x}_{k+1} = \underbrace{\mathbf{prox}_h}_{\text{projection}} \Big(\underbrace{\mathbf{prox}_g}_{\text{hard-thresholding}} (\mathbf{x}_k) \Big)}$$

THANK YOU FOR YOUR ATTENTION!