

AMATH 840: ADVANCED NUMERICAL
METHODS FOR COMPUTATIONAL AND DATA
SCIENCE

Winter 2023

Part 1: Compressive Sensing

1.3: ℓ_1 -Optimization Algorithms

Winter 2023

Relations between ℓ_1 -Optimization Models

Minimizing the Sum of Two Convex Functions

Proximal Operator

Minimizing the Sum of Two Convex Functions

Some Popular ℓ_1 -Optimization Algorithms

FISTA: A fast iterative shrinkage-thresholding algorithm

Nesterov's Second Method

spgl1 and Other Available Packages

Alternating Direction Method of Multipliers (ADMM)

Some Popular Algorithms for Compressive Sensing - Part 2

- Basis pursuit:

$$\min_{\mathbf{z} \in \mathbb{C}^n} \|\mathbf{z}\|_1 \quad \text{s.t.} \quad \mathbf{Az} = \mathbf{y}. \quad (\text{BP})$$

- Basis pursuit denoising:

$$\min_{\mathbf{z} \in \mathbb{C}^n} \|\mathbf{z}\|_1 \quad \text{s.t.} \quad \|\mathbf{Az} - \mathbf{y}\|_2 \leq \eta, \quad (\text{BP}_\eta)$$

or

$$\min_{\mathbf{z} \in \mathbb{C}^n} \frac{1}{2} \|\mathbf{Az} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{z}\|_1. \quad (\text{QP}_\lambda)$$

- Lasso:

$$\min_{\mathbf{z} \in \mathbb{C}^n} \frac{1}{2} \|\mathbf{Az} - \mathbf{y}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{z}\|_1 \leq \tau. \quad (\text{LS}_\tau)$$

We use the same numbers of Theorems, Lemmas, Propositions from “A Mathematical Introduction to Compressive Sensing”, by S. Foucart and H. Rauhut.

Relations between ℓ_1 -Optimization Models

Proposition 3.2. (Relations between (BP_η) , (QP_λ) , and (LS_τ)).

1. If \mathbf{z}_{qp} is a minimizer of (QP_λ) with $\lambda > 0$, then there exists $\sigma = \sigma_{\mathbf{z}_{qp}} \geq 0$ such that \mathbf{z}_{qp} is a minimizer of (BP_η) .
2. If \mathbf{z}_{bp} is a unique minimizer of (BP_η) with $\sigma \geq 0$, then there exists $\tau = \tau_{\mathbf{z}_{bp}} \geq 0$ such that \mathbf{z}_{bp} is a unique minimizer of (LS_τ) .
3. If \mathbf{z}_{ls} is a minimizer of (LS_τ) with $\tau > 0$, then there exists $\lambda = \lambda_{\mathbf{z}_{ls}} \geq 0$ such that \mathbf{z}_{ls} is a minimizer of (QP_λ) .

Proof Sketch.

- $(QP_\lambda \Rightarrow BP_\eta)$. Set $\sigma := \|A\mathbf{z}_{qp} - \mathbf{y}\|_2$.
- $(BP_\eta \Rightarrow LS_\tau)$. Set $\tau := \|\mathbf{z}_{bp}\|_1$.
- $(LS_\tau \Rightarrow QP_\lambda)$. See Theorem B.28 from “A Mathematical Introduction to Compressive Sensing”, by S. Foucart and H. Rauhut.

Relations between ℓ_1 -Optimization Models (cont'd)

With suitable values of η, λ, τ , the solutions of $BP_\eta, QP_\lambda, LS_\tau$ coincide.

- If A is orthogonal, a suggestion is $\lambda = \eta\sqrt{2\log(n)}$.¹
- In general, the relations among η, λ, τ cannot be known a priori.²
- If λ is large enough, the solution of QP_λ problem is $z_\lambda = 0$.

Theorem (BP vs QP_λ .)

Assume that $Aw = y$ has a solution. For each $\lambda > 0$, let z_λ be a minimizer of (QP_λ) . If the (BP) problem has a unique solution $z^\#$, then

$$\lim_{\lambda \rightarrow 0^+} z_\lambda = z^\#.$$

¹ *Atomic Decomposition by Basis Pursuit*, by Chen, Donoho, and Saunders, SIAM Review, 2001.

² *Probing the Pareto frontier for basis pursuit solutions*, by E. van den Berg and M. P. Friedlander, SIAM J. on Scientific Computing, 2008.

Relations between ℓ_1 -Optimization Models

Minimizing the Sum of Two Convex Functions

Proximal Operator

Minimizing the Sum of Two Convex Functions

Some Popular ℓ_1 -Optimization Algorithms

FISTA: A fast iterative shrinkage-thresholding algorithm

Nesterov's Second Method

spgl1 and Other Available Packages

Alternating Direction Method of Multipliers (ADMM)

Proximal Operator

Definition

Let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be a closed, proper, convex function, which means that its epigraph

$$\text{epi } f = \{(x, t) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq t\}$$

is a nonempty closed convex set.

- The **proximal operator** $\text{prox}_f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is defined as follows:

$$\text{prox}_f(\mathbf{x}_0) := \underset{\mathbf{x}}{\text{argmin}} \left\{ f(\mathbf{x}) + \frac{1}{2} \|\mathbf{x} - \mathbf{x}_0\|_2^2 \right\}, \quad \text{where } \mathbf{x}_0 \in \mathbb{R}^n.$$

- The proximal operator of the scaled function λf , where $\lambda > 0$,

$$\text{prox}_{\lambda f}(\mathbf{x}_0) := \underset{\mathbf{x}}{\text{argmin}} \left\{ \lambda f(\mathbf{x}) + \frac{1}{2} \|\mathbf{x} - \mathbf{x}_0\|_2^2 \right\}, \quad \text{where } \mathbf{x}_0 \in \mathbb{R}^n,$$

is also called the **proximal operator of f with parameter λ** .

Examples of Proximal Operators

- Example 1: The proximal operator of the ℓ_1 function is

$$\begin{aligned}\text{prox}_{\lambda\|\cdot\|_1}(\mathbf{x}_0) &:= \underset{\mathbf{x}}{\text{argmin}} \left\{ \lambda\|\mathbf{x}\|_1 + \frac{1}{2}\|\mathbf{x} - \mathbf{x}_0\|_2^2 \right\} \\ &= \text{sign}(\mathbf{x}_0) \max(|\mathbf{x}_0| - \lambda, 0). \quad (\text{element-wise})\end{aligned}$$

It is called a **soft-thresholding operator**.

- Example 2: The proximal operator of the ℓ_0 function is

$$\text{prox}_{\lambda\|\cdot\|_0}(\mathbf{x}_0) := \underset{\mathbf{x}}{\text{argmin}} \left\{ \lambda\|\mathbf{x}\|_0 + \frac{1}{2}\|\mathbf{x} - \mathbf{x}_0\|_2^2 \right\} = (\text{discuss in class}).$$

It is called a **hard-thresholding operator**.

- Example 3: The proximal operator of the indicator function is (discuss in class).

Minimizing the Sum of Two Convex Functions

- Consider the following nonsmooth convex optimization problem:

$$\min\{F(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^n\}$$

where

- $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is closed proper convex, continuously differentiable with Lipschitz continuous gradient $L(f)$:

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

- $g : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is closed proper convex, continuous function which is possibly nonsmooth, with inexpensive proximal operator $\text{prox}_g(\cdot)$.
 - The optimization problem is solvable, i.e., $\text{argmin } F \neq \emptyset$.
- In the remaining slides, we assume the functions f and g have those properties, unless stated otherwise.

Minimizing the Sum of Two Convex Functions (cont'd)

- Example: The ℓ_1 regularization problem (QP_λ), where

$$f(\mathbf{z}) = \frac{1}{2}\|A\mathbf{z} - \mathbf{y}\|^2, g(\mathbf{z}) = \lambda\|\mathbf{z}\|_1, \text{ and } L(f) = \lambda_{\max}(A^T A).$$

Relations between ℓ_1 -Optimization Models

Minimizing the Sum of Two Convex Functions

Proximal Operator

Minimizing the Sum of Two Convex Functions

Some Popular ℓ_1 -Optimization Algorithms

FISTA: A fast iterative shrinkage-thresholding algorithm

Nesterov's Second Method

spgl1 and Other Available Packages

Alternating Direction Method of Multipliers (ADMM)

FISTA: A fast iterative shrinkage-thresholding algorithm

- FISTA is an iterative shrinkage-thresholding algorithm (ISTA) with complexity result of $\mathcal{O}(1/k^2)$ (see Theorem 4.4 in ⁴) to solve

$$\min\{F(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^n\}.$$

FISTA with constant stepsize - $\min_{\mathbf{x} \in \mathbb{R}^n} (f(\mathbf{x}) + g(\mathbf{x}))$

Input: $L = L(f)$, a Lipschitz constant of ∇f , and final step K .

Step 0. $\mathbf{y}_1 = \mathbf{x}_0 \in \mathbb{R}^n$, $t_1 = 1$.

Step k. ($k \geq 1$) Compute

$$\begin{aligned}\mathbf{x}_k &= \operatorname{argmin}_{\mathbf{x}} \left\{ g(\mathbf{x}) + \frac{L}{2} \left\| \mathbf{x} - \left(\mathbf{y}_k - \frac{1}{L} \nabla f(\mathbf{y}_k) \right) \right\|^2 \right\} \\ &= \operatorname{prox}_{(1/L)g} \left(\mathbf{y}_k - \frac{1}{L} \nabla f(\mathbf{y}_k) \right), \\ t_{k+1} &= \frac{1 + \sqrt{1 + 4t_k^2}}{2}, \\ \mathbf{y}_{k+1} &= \mathbf{x}_k + \frac{t_k - 1}{t_{k+1}} (\mathbf{x}_k - \mathbf{x}_{k-1}).\end{aligned}$$

Output: \mathbf{x}_K

⁴ A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems, by Beck & Teboulle, SIAM J. Imaging Sciences, 2009.

FISTA - Another Version - $\min_{\mathbf{x} \in \mathbb{R}^n} (f(\mathbf{x}) + g(\mathbf{x}))$

Input: $L = L(f)$, a Lipschitz constant of ∇f , and final step K .

Step 0. Choose any $\mathbf{x}_1 = \mathbf{x}_0 \in \mathbb{R}^n$.

Step k. ($k \geq 1$) Compute

$$\mathbf{y} = \mathbf{x}_k + \frac{k-1}{k+2}(\mathbf{x}_k - \mathbf{x}_{k-1})$$
$$\mathbf{x}_{k+1} = \text{prox}_{t_{k+1}g} \left(\mathbf{y} - t_{k+1} \nabla f(\mathbf{y}) \right),$$

where step size $t_k = \frac{1}{L}, \forall k$ or is determined by line search.

Output: \mathbf{x}_K

5

⁵<http://www.seas.ucla.edu/~vandenbe/236C/lectures/fista.pdf>, by Boyd & Vandenberghe

FISTA: Examples

Using FISTA to solve

$$\min_{\mathbf{x} \in \mathbb{R}^n} \left(\frac{1}{2} \|\mathbf{Ax} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{x}\|_1 \right).$$

In this case, $f(\mathbf{x}) = \frac{1}{2} \|\mathbf{Ax} - \mathbf{b}\|_2^2$, $L(f) = \lambda_{\max}(A^T A)$, and $g(\mathbf{x}) = \lambda \|\mathbf{x}\|_1$.

FISTA to Solve - $\min_{\mathbf{x} \in \mathbb{R}^n} \left(\frac{1}{2} \|\mathbf{Ax} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{x}\|_1 \right)$

Input: Final step K .

Step 0. Choose any $\mathbf{x}_1 = \mathbf{x}_0 \in \mathbb{R}^n$.

Step k. ($k \geq 1$) Compute

$$\mathbf{y} = \mathbf{x}_k + \frac{k-1}{k+2} (\mathbf{x}_k - \mathbf{x}_{k-1})$$

$$\hat{\mathbf{y}} = \mathbf{y} - \frac{1}{L} A^T (A\mathbf{y} - \mathbf{b})$$

$$\mathbf{x}_{k+1} = \text{prox}_{(1/L)g}(\hat{\mathbf{y}}) = \text{sign}(\hat{\mathbf{y}}) \max(|\hat{\mathbf{y}}| - \frac{\lambda}{L}, 0),$$

where $L = \lambda_{\max}(A^T A)$.

Output: \mathbf{x}_K

Nesterov's Second Method

- Nesterov's second method is a gradient projection method with $(1/k^2)$ convergence rate.

Nesterov's Second Method

Input: $L = L(f)$, a Lipschitz constant of ∇f , and final step K .

Step 0. Choose any $\mathbf{x}_0 = \mathbf{z}_0 \in \mathbb{R}^n$.

Step k. ($k \geq 1$) Compute

$$\mathbf{y} = (1 - \theta_k)\mathbf{x}_{k-1} + \theta_k\mathbf{z}_{k-1}$$

$$\mathbf{z}_k = \text{prox}_{(t_k/\theta_k)g} \left(\mathbf{z}_{k-1} - \frac{t_k}{\theta_k} \nabla f(\mathbf{y}) \right)$$

$$\mathbf{x}_k = (1 - \theta_k)\mathbf{x}_{k-1} + \theta_k\mathbf{z}_k,$$

where $\theta_k = \frac{2}{k+1}$ and $t_k = \frac{1}{L}$, or one of the line search methods.

Output: \mathbf{x}_K

spgl1 and Other Available Packages

- spgl1⁸. Matlab and Python codes can be downloaded from <https://friedlander.io/spgl1/install>
- Python packages: scikit-learn package.
 - Link: https://scikit-learn.org/stable/modules/linear_model.html
 - Solve the (QP_λ) by coordinate descent method ⁹.

⁸*SPGL1: A solver for large-scale sparse reconstruction*, by Den Berg and Friedlander, 2007.

⁹*Regularization Path For Generalized linear Models by Coordinate Descent*, by Friedman, Hastie and Tibshirani.

Alternating Direction Method of Multipliers (ADMM)

Here we assume that f and g are convex, closed, proper and L_0 has a saddle point.

- Consider the following optimization problem:

$$\begin{aligned} & \text{minimize} && f(\mathbf{x}) + g(\mathbf{z}) \\ & \text{subject to} && \mathbf{Ax} + \mathbf{Bz} = \mathbf{c}. \end{aligned}$$

- The corresponding augmented Lagrangian is

$$L_\rho(\mathbf{x}, \mathbf{z}, \mathbf{y}) = f(\mathbf{x}) + g(\mathbf{z}) + \mathbf{y}^T(\mathbf{Ax} + \mathbf{Bz} - \mathbf{c}) + \frac{\rho}{2} \|\mathbf{Ax} + \mathbf{Bz} - \mathbf{c}\|_2^2$$

- ADMM algorithm:

$$\mathbf{x}_{k+1} := \underset{\mathbf{x}}{\operatorname{argmin}} L_\rho(\mathbf{x}, \mathbf{z}_k, \mathbf{y}_k) \quad (\mathbf{x}\text{-minimization})$$

$$\mathbf{z}_{k+1} := \underset{\mathbf{z}}{\operatorname{argmin}} L_\rho(\mathbf{x}_{k+1}, \mathbf{z}, \mathbf{y}_k) \quad (\mathbf{z}\text{-minimization})$$

$$\mathbf{y}_{k+1} := \mathbf{y}_k + \rho(\mathbf{Ax}_{k+1} + \mathbf{Bz}_{k+1} - \mathbf{c}) \quad (\text{dual update})$$

10 11 12

¹⁰ADMM is proposed by Gabay, Mercier, Glowinski, Marrocco in 1976.

¹¹*The Split Bregman Method for L1-Regularized Problems*, by Goldstein and Osher, SIAM J. Imaging Sciences, 2009.

¹²https://web.stanford.edu/class/ee364b/lectures/admm_slides.pdf

ADMM and Related Algorithms

- Under the stated assumptions, ADMM converges in the sense that
 - Iterates approach feasibility: $A\mathbf{x}_k + B\mathbf{z}_k - \mathbf{c} \rightarrow 0$
 - Objective approaches optimal value: $f(\mathbf{x}_k) + g(\mathbf{z}_k) \rightarrow p_*$
- Related algorithms:
 - operator splitting methods
 - proximal point algorithm
 - Bregman iterative methods

Example 1: Consider ADMM for

$$\min f(\mathbf{x}) \quad \text{subject to} \quad \mathbf{x} \in \mathbb{C}.$$

Answer:

- The ADMM form with $g(\mathbf{z}) = I_{\mathbb{C}}(\mathbf{z})$, the indicator function of set \mathbb{C} :

$$\begin{aligned} &\text{minimize} && f(\mathbf{x}) + g(\mathbf{z}) \\ &\text{subject to} && \mathbf{x} - \mathbf{z} = 0. \end{aligned}$$

- ADMM algorithm (discuss in class)

Example 2: Consider ADMM for

$$\min \frac{1}{2} \|A\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{x}\|_1.$$

Answer:

- The ADMM form with $g(\mathbf{z}) = \lambda \|\mathbf{z}\|_1$:

$$\text{minimize } \frac{1}{2} \|A\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{z}\|_1$$

$$\text{subject to } \mathbf{x} - \mathbf{z} = 0.$$

- ADMM algorithm (discuss in class)

Example 3: Consider ADMM for

$$\min \frac{1}{2} \|A\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \|C\mathbf{x} - \mathbf{d}\|_1.$$

Answer:

- The ADMM form with $g(\mathbf{z}) = \lambda \|\mathbf{z} - \mathbf{d}\|_1$:

$$\begin{aligned} & \text{minimize} && \frac{1}{2} \|A\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{z} - \mathbf{d}\|_1 \\ & \text{subject to} && C\mathbf{x} - \mathbf{z} = 0. \end{aligned}$$

- ADMM algorithm (discuss in class)

Example 4: Given a 2D noisy image f , consider ADMM for the TV denoising model:

$$\min_u \frac{\mu}{2} \|u - f\|_2^2 + \|\nabla_x u\|_1 + \|\nabla_y u\|_1.$$

Answer:

- The ADMM form:

$$\min_{u, d_x, d_y} \frac{\mu}{2} \|u - f\|_2^2 + \|d_x\|_1 + \|d_y\|_1$$

$$\text{subject to } d_x - \nabla_x u = 0 \quad \text{and} \quad d_y - \nabla_y u = 0.$$

- ADMM algorithm (discuss in class)

ℓ_1 -Algorithms for (QP_λ) Problem (cont'd)

Remarks:

- Global rate of convergence $\mathcal{O}(1/k^2)$ can be achieved, for example, with FISTA and Nesterov's 2nd method. ¹³
- The speed of some algorithms for ℓ_1 -minimization problems does not depend on the sparsity level s , such as the primal-dual algorithm \rightarrow Use ℓ_1 -minimization solvers for mildly large s .
- **Debiasing technique:** Suppose z_{sol} is the num. soln. of the (QP_λ) problem. Let $S := \text{supp}(z_{final})$ and solve

$$\min\{\|Az - y\|_2^2 : \text{supp}(z) \subset S\}.$$

¹³<http://www.seas.ucla.edu/~vandenbe/236C/lectures/fista.pdf>