A Dual Moving Balls Algorithm for a Class of Nonsmooth Convex Constrained Minimization

Marc Teboulle

School of Mathematical Sciences

Tel Aviv University

Joint work with

Ron Shefi

GDO 2019 – Games, Dynamics and Optimization
April 9 – 11, 2019, Babes-Bolyai University, Cluj-Napoca, Romania

The Problem

A First Order Method for Solving the Constrained NSO

$$(P) \qquad \varphi_* = \min\{\varphi(x) : g(x) \leq 0,\ x \in \mathbb{R}^n\},$$

- $\varphi : \mathbb{R}^n \to \mathbb{R}$ is a convex nonsmooth function
- $g : \mathbb{R}^n \to \mathbb{R}$ is convex $C^{1,1}_{L_g}$, i.e, L_g-Lipschitz continuous gradient on $\mathbb{R}^n$
- $\mathcal{F} := \{x \in \mathbb{R}^n :\ g(x) \leq 0\} \neq \emptyset$ the feasible set of (P).

Like in all FOM we assume that φ is “simple", that is *prox friendly*.

The Problem

A First Order Method for Solving the Constrained NSO

$$(P) \qquad \varphi_* = \min\{\varphi(x) : g(x) \leq 0, \; x \in \mathbb{R}^n\},$$

- $\varphi : \mathbb{R}^n \to \mathbb{R}$ is a convex nonsmooth function
- $g : \mathbb{R}^n \to \mathbb{R}$ is convex $C^{1,1}_{L_g}$, i.e, L_g-Lipschitz continuous gradient on $\mathbb{R}^n$
- $\mathcal{F} := \{x \in \mathbb{R}^n : \; g(x) \leq 0\} \neq \emptyset$ the feasible set of (P).

Like in all FOM we assume that φ is “simple", that is *prox friendly*.

Example:
Typical in various linear inverse problems: sparse recovery/ machine learning

$$\min\{\varphi(x) \equiv \mathrm{norm}(x) : \|Ax - b\|^2 \leq \delta, \; x \in \mathbb{E}\}.$$

Goal and Outline

Derive a simple $O(1/\varepsilon)$ first order algorithm to find an ε-optimal solution:

$$(P) \qquad \varphi_* = \min\{\varphi(x) \ : x \in \mathcal{F} \equiv \{x \in \mathbb{R}^n : g(x) \leq 0\}\}.$$

Using only data info and is Parameters Free.

Goal and Outline

Derive a simple $O(1/\varepsilon)$ first order algorithm to find an ε-optimal solution:

$$(P) \qquad \varphi_* = \min\{\varphi(x) \ : x \in \mathcal{F} \equiv \{x \in \mathbb{R}^n : g(x) \leq 0\}\}.$$

Using only data info and is Parameters Free.

- Underlying Idea of The New Method.
- Approach/Main Tools and Global Convergence Results.
- Numerical Example on Large Scale Sparse Recovery.

Don't We Have Other $O(1/\varepsilon)$ FOM for a Constrained NSO?

$$(P) \qquad \varphi_* = \min\{\varphi(x) \ : \ g(x) \leq 0, \ x \in \mathbb{R}^n\},$$

Fact: Best known rate with a first order method for (P): $O(1/\varepsilon)$.

Don't We Have Other $O(1/\varepsilon)$ FOM for a Constrained NSO?

$$(P) \qquad \varphi_* = \min\{\varphi(x) \; : \; g(x) \le 0, \; x \in \mathbb{R}^n\},$$

Fact: Best known rate with a first order method for (P): $O(1/\varepsilon)$.

ANSWER: Apparently, we don't ....!....

.....Without recourse to extra/unknowns parameters/heuristics...ect...........

Don't We Have Other $O(1/\varepsilon)$ FOM for a Constrained NSO?

$$(P) \qquad \varphi_* = \min\{\varphi(x) \; : \; g(x) \le 0, \; x \in \mathbb{R}^n\},$$

Fact: Best known rate with a first order method for (P): $O(1/\varepsilon)$.

ANSWER: Apparently, we don't ....!....

.....Without recourse to extra/unknowns parameters/heuristics...ect...........

- *Subgradient Projection/ Mirror Descent Type Methods* Slow convergent rate $O(1/\varepsilon^2)$...+ Need easy projection on *nonlinear constraint*. Bundle, same.
- *Fast Proximal-Gradient Methods* Great! Optimal rate $O(1/\sqrt{\varepsilon})$! But...to apply it one first must penalize the problem!...And we don't know the penalty parameter!
- *Smoothing Methods* Can tackle special forms of (P) with $O(1/\varepsilon)$ rate... But depends on smoothing and other parameters!
- *Lagrangian/ADM Methods* Even when they can... Need an unknown penalty parameter!... The complexity rate is $O(1/\varepsilon)$...But the constant depends on it ...! Large parameter $\Rightarrow$ very slow method!

Blanket Assumption on Convex NSO (P)

$$(P) \qquad \varphi_* = \min\{\varphi(x) : g(x) \leq 0,\ x \in \mathbb{R}^n\},\ \ [\varphi \text{ nonsmooth } g \in C^{1,1}_{L_g}].$$

Blanket Assumption A

A1 There exists an optimal solution for problem (P).

A2 Slater's condition holds: $\exists \hat{x} \in \mathbb{R}^n :\ g(\hat{x}) < 0$.

A3 For any $x \in \mathcal{F}$, $\mathbf{0} \notin \partial\varphi(x)$.

A1 and A2 are standard in convex programming. Warrant that $x^* \in \mathbb{R}^n$ is an optimal solution of (P) if and only if (KKT) optimality conditions hold, i.e.,

$$\text{[KKT-P]} \quad \exists\, \lambda^* \geq 0 \text{ such that } 0 \in \partial\varphi(x^*) + \lambda^* \nabla g(x^*);\ \lambda^* g(x^*) = 0, g(x^*) \leq 0.$$

A3 eliminates the trivial case: a feasible point as an unconstrained minimizer of $\varphi(\cdot)$.

The Approximation Model – Main Idea

Starting Idea: Approximate the feasible set by ***Moving Balls***.[Auslender-Shefi-Teboulle '10].

The Approximation Model – Main Idea

Starting Idea: Approximate the feasible set by ***Moving Balls***.[Auslender-Shefi-Teboulle '10].

- Exploit smoothness of g in the constraint. The descent Lemma gives for any $L \geq L_g$:

$$g(x) \leq g(y) + \langle \nabla g(y), x - y \rangle + \frac{L}{2}\|x - y\|^2, \quad \forall x, y \in \mathbb{R}^n.$$

The Approximation Model – Main Idea

Starting Idea: Approximate the feasible set by ***Moving Balls***.[Auslender-Shefi-Teboulle '10].

- Exploit smoothness of g in the constraint. The descent Lemma gives for any $L \geq L_g$:

$$g(x) \leq g(y) + \langle \nabla g(y), x - y \rangle + \frac{L}{2}\|x - y\|^2, \quad \forall x, y \in \mathbb{R}^n.$$

- **Algebra Time...The Descent Lemma Reads:**

$$\frac{2}{L} g(x) \leq \|x - c(y)\|^2 - \rho^2(y),$$

where

$$\begin{aligned} c(y) &:= y - (1/L)\nabla g(y), \\ \rho^2(y) &:= \frac{1}{L^2}\|\nabla g(y)\|^2 - \frac{2}{L} g(y). \end{aligned}$$

Leads to the following approximation of problem (P)...

The Approximate Convex Model

Fix any $y \in \mathcal{F}$.

Define the ball centered at $c(y)$ with radius $\rho(y)$

$$\mathcal{B}(y) := \{x \in \mathbb{R}^n : \; \|x - c(y)\|^2 \leq \rho^2(y)\}.$$

The Approximated Convex Problem $P(y)$

For each $y \in \mathcal{F}$ minimizes the nonsmooth objective over the ball $B(y)$:

$$(P(y)) \qquad \begin{array}{ll} \min & \varphi(x) \\ \text{subject to} & x \in B(y). \end{array}$$

Problem $P(y)$ is *a natural approximation* of problem (P).
This is justified by the following properties which also lead to the algorithm.

Fix any $y \in \mathcal{F}$.

Proposition 1 - [Approximation of (P)]

1. (i) $B(y)$ is a nonempty, compact convex set with $B(y) \subseteq \mathcal{F}$.
2. (ii) Slater's condition holds for problem $P(y)$.

Basic Properties of $P(y)$: $\min\{\varphi(x) : x \in B(y)\}$

Fix any $y \in \mathcal{F}$.

Proposition 1 - [Approximation of (P)]

(i) $B(y)$ is a nonempty, compact convex set with $B(y) \subseteq \mathcal{F}$.

(ii) Slater's condition holds for problem $P(y)$.

Proposition 2- [Fixed Point] Let $x(y)$ be an optimal solution for problem $P(y)$.

If $y \in x(y)$, then y is a solution for problem (P).

Basic Properties of $P(y)$: $\min\{\varphi(x) : x \in B(y)\}$

Fix any $y \in \mathcal{F}$.

Proposition 1 - [Approximation of (P)]

- (i) $B(y)$ is a nonempty, compact convex set with $B(y) \subseteq \mathcal{F}$.
- (ii) Slater's condition holds for problem $P(y)$.

Proposition 2- [Fixed Point] Let $x(y)$ be an optimal solution for problem $P(y)$.

If $y \in x(y)$, then y is a solution for problem (P).

Basic Scheme Generate a sequence of feasible (interior) pts by minimizing φ over a sequence of moving balls.

$$x^0 \in \mathcal{F}, \qquad x^k \in \operatorname{argmin}\left\{\varphi(x) : \|x - c(x^{k-1})\|^2 \le \rho^2(x^{k-1})\right\}, \ k \ge 1.$$

How to implement this?

Basic Properties of $P(y)$: $\min\{\varphi(x) : x \in B(y)\}$

Fix any $y \in \mathcal{F}$.

Proposition 1 - [Approximation of (P)]

(i) $B(y)$ is a nonempty, compact convex set with $B(y) \subseteq \mathcal{F}$.

(ii) Slater's condition holds for problem $P(y)$.

Proposition 2- [Fixed Point] Let $x(y)$ be an optimal solution for problem $P(y)$.

If $y \in x(y)$, then y is a solution for problem (P).

Basic Scheme Generate a sequence of feasible (interior) pts by minimizing φ over a sequence of moving balls.

$$x^0 \in \mathcal{F}, \qquad x^k \in \operatorname{argmin}\left\{\varphi(x) : \|x - c(x^{k-1})\|^2 \leq \rho^2(x^{k-1})\right\}, \; k \geq 1.$$

How to implement this?

Our approach: Solve $P(y)$ via its dual!

A Dual Problem for P(y)

Fix any $y \in \mathcal{F}$

$$P(y) \qquad \min\{\varphi(x) : \ \|x - c(y)\|^2 \leq \rho^2(y), \ x \in \mathbb{R}^n\}.$$

A Lagrangian dual for $P(y)$ is *one dimensional* convex problem in λ

$$D(y) \qquad \sup\{q(\lambda; y) : \ \lambda \geq 0\} \equiv \sup\{q(\lambda; y) : \ \lambda > 0\}.^{1}$$

with

$$q(\lambda; y) := -\frac{\lambda}{2}\rho(y)^2 + \min_{x \in \mathbb{R}^n}\{\varphi(x) + \frac{\lambda}{2}\|x - c(y)\|^2\}.$$

The dual objective is *one dimensional*...with nice properties..!

[1] Last equality can be proven thanks to closedness of q.

The Dual Objective is Very Nice!

The dual objective is a *one dimensional* concave function in λ:

$$\lambda \to q(\lambda; y) = \underbrace{\min_{x \in \mathbb{R}^n} \{\varphi(x) + \frac{\lambda}{2}\|x - c(y)\|^2\}}_{M_\lambda^\varphi(c(y))} - \frac{\lambda}{2}\rho(y)^2$$

The dual variable is nothing else but **the proximal parameter** in the Moreau's envelope of the nonsmooth $\varphi(\cdot)$:

$$q(\lambda; y) \equiv M_\lambda^\varphi(c(y)) - \frac{\lambda}{2}\rho^2(y).$$

The Dual Objective is Very Nice!

The dual objective is a *one dimensional* concave function in λ:

$$\lambda \to q(\lambda; y) = \underbrace{\min_{x \in \mathbb{R}^n} \{\varphi(x) + \frac{\lambda}{2}\|x - c(y)\|^2\}}_{M_\lambda^\varphi(c(y))} - \frac{\lambda}{2}\rho(y)^2$$

The dual variable is nothing else but **the proximal parameter** in the Moreau's envelope of the nonsmooth $\varphi(\cdot)$:

$$q(\lambda; y) \equiv M_\lambda^\varphi(c(y)) - \frac{\lambda}{2}\rho^2(y).$$

- Properties of proximal maps and their envelopes are very well known and useful.....
- But here we are interested in the properties of the proximal envelope $M_\lambda^\varphi(u)$ as a **function of the parameter** $\lambda > 0$

$$\lambda \to M_\lambda^\varphi(u), \qquad \text{when } \underline{u \in \mathbb{R}^d \text{ is fixed}}.$$

Proximal Maps/Envelopes as Function of Proximal Parameter

Let $h : \mathbb{R}^d \to (-\infty, +\infty]$ be a closed proper convex function. For any $u \in \mathbb{R}^d$ and any $t > 0$, the *proximal map* of h and its *proximal envelope* are defined respectively by:

$$\text{prox}_t^h(u) = \text{argmin}_{z \in \mathbb{R}^d} \left\{ h(z) + \frac{t}{2} \|z - u\|^2 \right\}$$

$$M_t^h(u) = \min_{z \in \mathbb{R}^d} \left\{ h(z) + \frac{t}{2} \|z - u\|^2 \right\}$$

Proximal Maps/Envelopes as Function of Proximal Parameter

Let $h : \mathbb{R}^d \to (-\infty, +\infty]$ be a closed proper convex function. For any $u \in \mathbb{R}^d$ and any $t > 0$, the *proximal map* of h and its *proximal envelope* are defined respectively by:

$$\text{prox}_t^h(u) = \text{argmin}_{z \in \mathbb{R}^d} \left\{ h(z) + \frac{t}{2}\|z - u\|^2 \right\}$$

$$M_t^h(u) = \min_{z \in \mathbb{R}^d} \left\{ h(z) + \frac{t}{2}\|z - u\|^2 \right\}$$

Proposition For any $u \in \mathbb{R}^d$, the following properties hold for $t \to M_t^h(u)$:

- (i) The function $t \to M_t^h(u)$ is concave and $C^1(0, \infty)$ with derivative

$$\frac{d}{dt} M_t^h(u) = \frac{1}{2}\|\text{prox}_t^h(u) - u\|^2.$$

- (ii) For any $u \in \text{dom}\, h$, $\lim_{t \to \infty} M_t^h(u) = h(u)$ and $\lim_{t \to \infty} \text{prox}_t^h(u) = u$.
- (iii) $\lim_{t \to 0^+} M_t^h(u) = -h^*(0)$.
- (iv) $\lim_{t \to 0^+} \text{prox}_t^h(u) = \text{argmin}\{h(u) : u \in \mathbb{R}^d\} = \partial h^*(0)$

Thanks to this, we can derive useful properties for the dual function $q(\lambda; y)$.

Properties of The Dual Objective of $P(y)$

Fix any $y \in \mathcal{F}$,and let $\psi : (0, \infty) \to \mathbb{R}$

$$\psi(\lambda) := q(\lambda; y) = M_\lambda^\varphi(c(y)) - \frac{\lambda}{2}\rho^2(y).$$

Apply previous proposition to get the following

Properties of the dual function $\psi(\lambda)$.

(i) ψ is a concave, $C^1(0, \infty)$, with derivative

$$\psi'(\lambda) = \frac{1}{2}\left\{\|\text{prox}_\lambda^\varphi(c(y)) - c(y)\|^2 - \rho^2(y)\right\}.$$

(ii) An optimal solution of the dual problem $\bar{\lambda} > 0$ solves the scalar equation

$$\psi'(\lambda) = 0.$$

Using these, we are ready to define the primal-dual algorithm for solving $P(y)$.

The Algorithm: DUal Moving Ball Algorithm (DUMBA)

DUMBA

Let $x^0 \in \mathcal{F}$, and for $k = 1, 2, \ldots$, generate $x^k \in \mathcal{F}$ and $\lambda_k \in (0, \infty)$ via the iterations:

Step 1. Compute

$$c(x^{k-1}) = x^{k-1} - (1/L)\nabla g(x^{k-1}), \qquad \rho(x^{k-1})^2 = (1/L^2)\|\nabla g(x^{k-1})\|^2 - (2/L)g(x^{k-1}).$$

Step 2. Find a positive root λ for the scalar equation

$$\|x(\lambda) - c(x^{k-1})\|^2 = \rho^2(x^{k-1}),$$

where $x(\lambda) := \text{prox}_{\lambda}^{\varphi}\left(c(x^{k-1})\right)$, and set $\lambda_k = \lambda$.

Step 3. Update

$$x^k = \text{prox}_{\lambda_k}^{\varphi}\left(c(x^{k-1})\right).$$

The Algorithm: DUal Moving Ball Algorithm (DUMBA)

DUMBA

Let $x^0 \in \mathcal{F}$, and for $k = 1, 2, \ldots$, generate $x^k \in \mathcal{F}$ and $\lambda_k \in (0, \infty)$ via the iterations:

Step 1. Compute

$$c(x^{k-1}) = x^{k-1} - (1/L)\nabla g(x^{k-1}), \qquad \rho(x^{k-1})^2 = (1/L^2)\|\nabla g(x^{k-1})\|^2 - (2/L)g(x^{k-1}).$$

Step 2. Find a positive root λ for the scalar equation

$$\|x(\lambda) - c(x^{k-1})\|^2 = \rho^2(x^{k-1}),$$

where $x(\lambda) := \text{prox}_{\lambda}^{\varphi}\left(c(x^{k-1})\right)$, and set $\lambda_k = \lambda$.

Step 3. Update

$$x^k = \text{prox}_{\lambda_k}^{\varphi}\left(c(x^{k-1})\right).$$

The main computational step in DUMBA consists of:

1. computing the proximal map of φ at a given c (like in all prox-gradient methods).
2. Solving a scalar equation. Price to pay to handle nonlinear constraint!

Convergence Result I – Point Convergence

Theorem 1 [Pointwise Convergence]

Let $\{x^k\}$ be the sequence generated by DUMBA. Then,

(i) the sequence of function values $\{\varphi(x^k)\}$ is monotonically decreasing,

(ii) the sequence $\{x^k\}$ is bounded and converges to an optimal solution of problem (P).

Convergence Result II – Complexity

Theorem 2 (Global Rate in Function Values)
Let $\{x^k\} \in \mathcal{F}$ and $\lambda_k \in (0, \infty)$ be the primal-dual sequences generated by DUMBA, and let x^* be an optimal solution of (P). Then, for all $k \geq 1$,

- there exists a positive constant C such $\lambda_k \leq C$,
- we have
$$\varphi(x^k) - \varphi(x^*) \leq \frac{C\|x^0 - x^*\|^2}{k}.$$

Note: The positive constant C depends on the problem's data.

Convergence Result II – Complexity

Theorem 2 (Global Rate in Function Values)
Let $\{x^k\} \in \mathcal{F}$ and $\lambda_k \in (0, \infty)$ be the primal-dual sequences generated by DUMBA, and let x^* be an optimal solution of (P). Then, for all $k \geq 1$,

- there exists a positive constant C such $\lambda_k \leq C$,
- we have

$$\varphi(x^k) - \varphi(x^*) \leq \frac{C\|x^0 - x^*\|^2}{k}.$$

Note: The positive constant C depends on the problem's data.

Open Question: Can we determine C explicitly?

One Answer: for a Special, but Important Class of Problems

$$\min\{\varphi(x) : \|Ax - b\| \leq \delta, \ x \in \mathbb{E}\}.$$

- The objective φ is assumed Lipschitz continuous with known constant L_φ.
- $AA^T = I$ (i.e., a restricted isometry).

Covers usual regularization of linear inverse problems with any norm in the objective.

One Answer: for a Special, but Important Class of Problems

$$\min\{\varphi(x) : \|Ax - b\| \leq \delta, \ x \in \mathbb{E}\}.$$

- The objective φ is assumed Lipschitz continuous with known constant L_φ.
- $AA^T = I$ (i.e., a restricted isometry).

Covers usual regularization of linear inverse problems with any norm in the objective.

Answer for that class. The complexity constant is explicitly given by:

$$C = \frac{L_\varphi}{\delta}.$$

One Answer: for a Special, but Important Class of Problems

$$\min\{\varphi(x) : \|Ax - b\| \le \delta,\ x \in \mathbb{E}\}.$$

- The objective φ is assumed Lipschitz continuous with known constant L_φ.
- $AA^T = I$ (i.e., a restricted isometry).

Covers usual regularization of linear inverse problems with any norm in the objective.

Answer for that class. The complexity constant is explicitly given by:

$$C = \frac{L_\varphi}{\delta}.$$

- DUMBA can be very slow whenever δ is too small...Not a surprise!
- Obviously not designed to solve *equality* constrained problems!

One Answer: for a Special, but Important Class of Problems

$$\min\{\varphi(x) : \|Ax - b\| \leq \delta, \; x \in \mathbb{E}\}.$$

- The objective φ is assumed Lipschitz continuous with known constant L_φ.
- $AA^T = I$ (i.e., a restricted isometry).

Covers usual regularization of linear inverse problems with any norm in the objective.

Answer for that class. The complexity constant is explicitly given by:

$$C = \frac{L_\varphi}{\delta}.$$

- DUMBA can be very slow whenever δ is too small...Not a surprise!
- Obviously not designed to solve *equality* constrained problems!
- **The positive side:** becomes fast and useful for a **large perturbation** δ!
- Numerical experiments confirm the theory.

Examples

- We illustrate the main step of DUMBA on several well known convex models arising in various applications: machine learning, signal processing, etc..

$$\text{find } \lambda > 0 \text{ that solves } \|\text{prox}_\lambda^\varphi(c) - c\|^2 = \rho^2.$$

- In all the examples below the objective function φ will be a norm on an appropriate Euclidean space.

  In that case, A3 eliminates the trivial optimal solution $x^* = 0$ in problem (P), and translates to

$$g(0) > 0 \implies \|c\| > \rho,$$

  which is exactly what is needed to warrant solution of the scalar equation.

Typical Examples in Sparse Recovery

Compute $\text{prox}_\lambda^\varphi(c)$ and find $\lambda > 0$ that solves $\|\text{prox}_\lambda^\varphi(c) - c\|^2 = \rho^2$.

All cases below are with φ 'prox friendly", i.e., explicit formula.

$$\varphi(x) = \|x\|_2 \quad - \quad \textbf{Euclidean norm}$$

$$\varphi(x) = \|x\|_1 \quad - \quad l_1 - \textbf{norm}$$

$$\varphi(x) = \sum_{g \in \mathcal{G}} \|x_g\|_2 \quad - \quad \textbf{Group lasso mixed norm } l_1/l_2,\ \mathcal{G} \text{ partition } \{1, \ldots, g\}$$

$$\varphi(x) = \|X\|_* \quad - \quad \textbf{Trace norm } X \in \mathbb{R}^{n \times n}.$$

- First example admits a closed formula for $\lambda = 1/\rho$.
- Remaining examples λ solves a scalar equation of similar type, e.g., for l_1:

$$\sum_{i=1}^{n} \min\left\{|c_i|^2, \frac{1}{\lambda^2}\right\} = \rho^2.$$

Efficient procedures in $O(n)$ [Bruker, 1984].

Numerical Examples

- We tested DUMBA on the BPDN, a central model for sparse recovery.

$$\text{(BPDN)} \qquad \begin{array}{ll} \text{minimize} & \|x\|_1 \\ \text{subject to} & \|Ax - b\|_2^2 \leq \delta^2, \; x \in \mathbb{R}^n, \end{array}$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, δ^2 is the noise power estimates.

Numerical Examples

- We tested DUMBA on the BPDN, a central model for sparse recovery.

$$\text{(BPDN)} \quad \begin{array}{ll} \text{minimize} & \|x\|_1 \\ \text{subject to} & \|Ax - b\|_2^2 \le \delta^2, \ x \in \mathbb{R}^n, \end{array}$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, δ^2 is the noise power estimates.

- Comparison vs **NESTA** [Becker et al. 2010] $\equiv$ Smoothing + Optimal Gradient.

  Complexity $O(1/k^2)$ **.. But for the "smoothed" objective** φ_μ.

  • Already includes sets of extensive experiments and comparison with other state-of-the-art methods for solving this class of problems.

- We tested DUMBA on the BPDN, a central model for sparse recovery.

$$\text{(BPDN)} \quad \begin{array}{ll} \text{minimize} & \|x\|_1 \\ \text{subject to} & \|Ax - b\|_2^2 \le \delta^2, \; x \in \mathbb{R}^n, \end{array}$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, δ^2 is the noise power estimates.

- Comparison vs **NESTA** [Becker et al. 2010] $\equiv$ Smoothing + Optimal Gradient.

  Complexity $O(1/k^2)$ **.. But for the "smoothed" objective** φ_μ.

  • Already includes sets of extensive experiments and comparison with other state-of-the-art methods for solving this class of problems.

- The strength of DUMBA-L1 [specialized to this problem's model]

  • Complexity of $O(1/k)$ **.. But for the "original" objective** φ.

  • Parameters free – No smoothing or other parameters to guess or tune.

  • Allows – **in fact dedicated to!**– for efficiently handling a large error δ.

The DUMBA -L_1 Algorithm Versus NESTA

Tested on random problems with $n = 262,144; \; m = n/8; \; s = m/5.$

Experimental Setup from NESTA: Low and High Dynamic Ranges

Dynamic range d**:** is a measure of the ratio between the largest and smallest magnitudes of the non-zero coefficients of the unknown signal.

- High dynamic range: useful in applications requiring detection and recovery of signals with small amplitudes obscures by large ones.
- Low dynamic range: useful for problems with large errors δ.

The DUMBA -L_1 Algorithm Versus NESTA

Tested on random problems with $n = 262,144;\ m = n/8;\ s = m/5.$

Experimental Setup from NESTA: Low and High Dynamic Ranges

Dynamic range d**:** is a measure of the ratio between the largest and smallest magnitudes of the non-zero coefficients of the unknown signal.

- High dynamic range: useful in applications requiring detection and recovery of signals with small amplitudes obscures by large ones.
- Low dynamic range: useful for problems with large errors δ.

As expected from the theory, DUMBA-L1:

- is effective in obtaining good accuracy of approximate solution with sparse signals at **low dynamic range** $\Longleftrightarrow$ **large bounding error**.
- is less efficient in reaching an extremely high accuracy when in the high dynamic range setting. [..Like any other methods..!].
- Yet, remains of comparable quality and speed vs available state of the art schemes.

Numerical Experiments Setup

$$(BPDN) \qquad \text{minimize } \{\|x\|_1 : \|Ax - b\|_2^2 \le \delta^2,\ x \in \mathbb{R}^n\} \tag{1}$$

- $x_s \in \mathbb{R}^n$ is s-sparse signal.
- $A \in \mathbb{R}^{m \times n}$ is a randomly subsampled discrete cosine transform, $AA^T = I$.
- $n = 262,144$, $m = n/8$, and $s = m/5$.
- $b = Ax_s + e$ with $e \sim N(0, \sigma)$. Noise level $\sigma = 0.1$.
- $\delta = \sqrt{m + 2\sqrt{2m}}\sigma$,

Following the experiment setup of NESTA,

$$x[i] = \mathbb{I}(i \in \Lambda)\eta_1[i]10^{\alpha\eta_2[i]}, \tag{2}$$

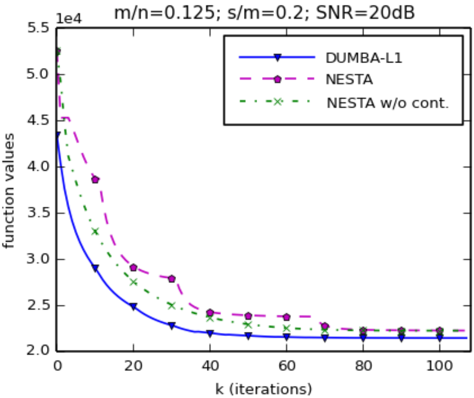
Λ - choosing s indices from the set $[n]$.
$\eta_1[i], i \in \Lambda$ - iid Bernoulli random variables.
$\eta_2[i], i \in \Lambda$ - iid Uniformly distributed random variables in $[0, 1]$.

The signal x_s created in this manner have a dynamic range d dB, where $\alpha = d/20$.

low dynamic range $d = 20dB$
high dynamic range $d = 40, 60, 80, 100dB$

Comparison with NESTA: Low Dynamic $d = 20$dB

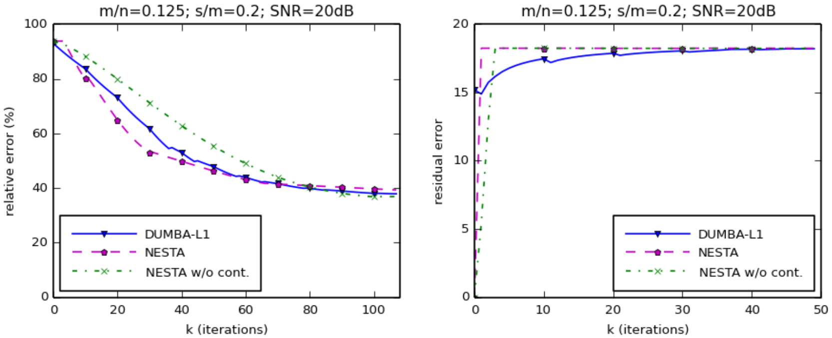
Figure: DUMBA-L1 and NESTA with and without continuation SNR=20dB. Function values vs # iterations



- **NETSA Needs**: $T = 5, \mu_f = 0.02, \texttt{tol} = 10^{-5}$.
- **DUMBA-L1**: Only stopping tolerance parameter $\texttt{tol}$=10^{-5}.

Comparison with NESTA: Low Dynamic $d = 20$dB

Figure: DUMBA-L1 and NESTA with/without continuation Relative error and residual error vs # iterations



Comparison with NESTA: Varying High Dynamic Range

Table: Comparison of accuracy: DUMBA-L1 with continuation and NESTA - N_A= number of calls to A, A^T

dB	Method	N_A	$\frac{\|x-x_s\|_2}{\|x_s\|_2}$	$\|x\|_1$	$\|Ax-b\|_2$
40	NESTA	458	0.06610	137,745.0546	18.2425
	DUMBA-L1	410	0.06601	136,952.0288	18.2421
60	NESTA	581	0.00979	940,690.6773	18.2419
	DUMBA-L1	602	0.00979	939,900.5124	18.2419
80	NESTA	575	0.00163	7,047,085.3632	18.2409
	DUMBA-L1	614	0.00162	7,046,316.2648	18.2349
100	NESTA	604	0.00035	56,155,527.2231	18.2400
	DUMBA-L1	720	0.00032	56,154,276.5952	18.2376

- The stopping criteria used was

$$\|\hat{x}^k\|_1 \leq \|x_{\rm NESTA}\|_1 \quad \text{and} \quad \|\hat{x}^k - x_s\|_2/\|x_s\| \leq \|x_{\rm NESTA} - x_s\|_2/\|x_s\| \quad \text{and}$$
$$\|A\hat{x}^k - b\|_2 \leq \delta.$$

- **NESTA**: $\texttt{tol}_{\rm NESTA} = 10^{-6}$
- **DUMBA-L1**: $\texttt{tol}_f = 10^{-6}$ and $T = 7$ continuation steps.

Thank you for your attention!