Double Smoothing technique for Convex Optimization Problems with Linear Constraints

O. Devolder (F.R.S.-FNRS Research Fellow),
F. Glineur and Y. Nesterov

Center for Operations Research and Econometrics (CORE),
Université catholique de Louvain (UCL)

ORBEL 25, Gent, February 11, 2011

Outline

1. Studied problem class.
2. Dual Approach.
3. Double Regularization.
4. Solving the dual problem.
5. Reconstruction of a primal solution.
6. Applications and Further Research.

A Class of Convex Optimization problems

$$P^* = \min_{u \in S} J(u)$$

$$\mathcal{A}u \in T$$

where:

- $J : U \to \mathbb{R}$ is a closed convex function
- S is a bounded, closed convex set in U
- U is an Hilbert space (possibly **infinite-dimensional**)
- T is a bounded, closed, convex set in V^*
- V is a **finite-dimensional** Hilbert space.

Furthermore S and T are simple i.e. that projections on these sets can be computed easily.

Two Important Assumptions

1. Without the linear constraint, the problem:

$$\min_{u \in S} J(u)$$

   is easy.
   Consequence: A natural approach is to dualize the linear constraint.

2. $$\dim V << \dim U$$

   Consequence: We want a purely dual algorithmic scheme, generating iterates only in the small-dimensional space V.

Outline

1. Studied problem class.
2. Dual Approach.
3. Double Regularization.
4. Solving the dual problem.
5. Reconstruction of a primal solution.
6. Applications and Further Research.

Primal/Dual Problems

- Primal Problem:

$$P^* = \min_{u \in S} \{J(u) + \max_{z \in V} [\langle \mathcal{A}u, z\rangle - \sigma_T(z)]\}$$

- Dual Problem:

$$D^* = \max_{z \in V} \{-\sigma_T(z) + \min_{u \in S} [J(u) + \langle \mathcal{A}u, z\rangle]\}$$

where $\sigma_T(z) = \sup_{x \in T} \langle x, z\rangle =$ support function of T.

Why a Dual Approach ?

Advantages of the dual problem:

- The dual is an unconstrained optimization problem in the small-dimensional space V.
- The subproblems defining $\phi(z) = \max_{u \in S}[-J(u) - \langle \mathcal{A}u, z \rangle]$ and $\sigma_T(z) = \max_{x \in T}\langle x, z \rangle$ can be solved easily (or even in closed-form) for any $z \in V$.
  $\Rightarrow$ We can solve the dual problem (in minimization form):
  $$-D^* = \Theta^* = \inf_{z \in V} [\sigma_T(z) + \phi(z) := \theta(z)]$$
  by a first-order method.
- Strong Duality holds i.e : $P^* = D^*$ under mild assumptions.

But the dual function can be non-differentiable...

$$-D^* = \theta^* = \min_{z \in V} \theta(z) = \min_{z \in V} \sigma_T(z) + \phi(z)$$

σ_T and ϕ can be **non-differentiable**:

- $\partial \sigma_T(z) = \{\tilde{x} \in T : \langle \tilde{x}, z \rangle = \sigma_T(z)\}$
- $\partial \phi(z) =$
  $\{-\mathcal{A}\tilde{u}$ for any optimal solution $\tilde{u}$ of the problem defining $\phi(z)\}$.

Conclusion:

We have to solve a non-smooth convex optimization problem.

How to solve a non-smooth convex problem ?

- **The classical approach: subgradient-type scheme.**
  Advantage : Can be applied directly on the dual objective function without any regularization
  Disadvantage: Slow Convergence
  $$\theta(z_k) \to \theta^* \text{ in } O\left(\frac{1}{\epsilon^2}\right).$$
- **The smoothing approach.**
  We modify the dual objective function in order to be able to apply more efficient scheme of smooth convex optimization.
  Advantage : Faster convergence, we will obtain a scheme such that
  $$\theta(z_k) \to \theta^* \text{ in } O\left(\frac{1}{\epsilon}\ln\left(\frac{1}{\epsilon}\right)\right).$$
  Disadvantage : We have to modify the dual objective function with some regularizations.

Outline

1. Studied problem class.
2. Dual Approach.
3. Double Regularization.
4. Solving the dual problem.
5. Reconstruction of a primal solution.
6. Applications and Further Research.

Double Regularization of the dual objective function

In order to

- be able to solve efficiently the dual problem
- to be able to obtain a nearly optimal and feasible primal solution from a nearly optimal dual solution

we will modify the dual objective function with two regularizations:

1. A first regularization that makes the dual objective function, **smooth with a Lipschitz-continuous gradient**
2. A second regularization that makes the dual objective function **strongly convex.**

First regularization

Why?
In order to obtain a **smooth dual objective function** with Lipschitz-continuous gradient.
$\Rightarrow$ We can apply efficient schemes of smooth convex optimization.
How? Let $\rho, \mu > 0$,
we modify the dual objective function:

$$\begin{aligned} \theta(z) &= \sigma_T(z) + \phi(z) \\ &= \sup_{x \in T} \langle x, z \rangle + \sup_{u \in S} [-J(u) - \langle \mathcal{A}u, z \rangle] \end{aligned}$$

in

$$\begin{aligned} \theta_{\rho,\mu}(z) &= \sigma_{\rho,T}(z) + \phi_{\mu}(z) \\ &= \sup_{x \in T} \{ \langle x, z \rangle - \frac{\rho}{2} \|x\|^2_{V^*} \} \\ &\quad + \sup_{u \in S} [-J(u) - \langle \mathcal{A}u, z \rangle - \frac{\mu}{2} \|u\|^2_U]. \end{aligned}$$

First regularization: $\theta_{\rho,\mu}$ is differentiable

With the addition of the strongly concave functions $-\frac{\rho}{2}\|x\|^2_{V^*}$ and $-\frac{\mu}{2}\|u\|^2_U$, the optimization problems defining $\sigma_{\rho,T}(z)$ and $\phi_\mu(z)$ have both only one optimal solution:

$$x_{\rho,z} = \arg\max_{x \in T}\{\langle x, z\rangle - \frac{\rho}{2}\|x\|^2_{V^*}\}$$

$$u_{\mu,z} = \arg\max_{u \in U}[-J(u) - \langle \mathcal{A}u, z\rangle - \frac{\mu}{2}\|u\|^2_U].$$

The function $\theta_{\rho,\mu}$ is therefore differentiable with gradient:

$$\nabla\theta_{\rho,\mu}(z) = x_{\rho,z} - \mathcal{A}u_{\mu,z}.$$

First regularization: Further properties of $\theta_{\rho,\mu}$

- $\theta_{\rho,\mu}$ has a Lipschitz-continuous gradient with constant

$$L_{\rho,\mu} = \frac{1}{\rho} + \frac{\|\mathcal{A}\|^2}{\mu}$$

- $\theta_{\rho,\mu}$ is a good approximation of θ with absolute accuracy bound depending on ρ and μ:

$$\theta_{\rho,\mu}(z) \leq \theta(z) \leq \theta_{\rho,\mu}(z) + \rho D_T + \mu D_S \quad \forall z \in V$$

where

$$\begin{aligned} D_T &= \max_{x \in T} \frac{1}{2} \|x\|_{V^*}^2 \\ D_S &= \max_{u \in S} \frac{1}{2} \|u\|_U^2 . \end{aligned}$$

Second regularization

Why?

We want not only

- to solve the dual problem

but also

- to reconstruct from the obtained nearly optimal dual solution, a nearly optimal and feasible primal solution.

Reconstruction of a primal solution

For a given dual iterate z_k, if we consider

$$u_k = u_{\mu,z_k},$$

the unique optimal solution of the problem defining $\phi_\mu(z_k)$, we have:

- $$J(u_k) \leq P^* + |\theta(z_k) - \theta^*| + Cst\,\|\nabla\theta_{\rho,\mu}(z_k)\|_{V^*} + 2\rho D_T + 2\mu D_S$$

- $$\|\mathcal{A}u_k - x_{\rho,z_k}\|_{V^*} = \|\nabla\theta_{\rho,\mu}(z_k)\|_{V^*}$$

  where $x_{\rho,z_k} \in C$.

The quality of this primal solution depends not only on the convergence rate of $\theta(z_k)$ to θ^* but also on the convergence rate of $\|\nabla\theta_{\rho,\mu}(z_k)\|_{V^*}$ to 0.

Convexity and Lipschitz-continuity of the gradient are not enough

If we apply the optimal scheme for $F^{1,1}_{L_{\rho,\mu}}(V)$ to our function $\theta_{\rho,\mu} \in F^{1,1}_{L_{\rho,\mu}}(V)$ with a good choice for ρ and μ, we have

$$\theta(z_k) - \theta^* \to 0 \text{ in } O\left(\frac{1}{\epsilon}\right)$$

but

$$\left\|\nabla\theta_{\rho,\mu}(z_k)\right\|_{V^*} \to 0 \text{ in } O\left(\frac{1}{\epsilon^2}\right).$$

Therefore if the dual objective function is only convex and smooth with a Lipschitz-continuous gradient, we have a convergence rate in $O\left(\frac{1}{\epsilon^2}\right)$ for the primal sequence.
This is not better than with the subgradient scheme!

Is the smoothing approach useless ?

No!!!

If we add a strongly convex term to $\theta_{\rho,\mu}$, we obtain a dual objective function: $\theta_{\rho,\mu,\kappa}(z) = \theta_{\rho,\mu}(z) + \frac{\kappa}{2}\|z\|_V^2 \in S^{1,1}_{\kappa,L_{\rho,\mu}+\kappa}(V)$.

Now, applying the optimal scheme for $S^{1,1}_{\kappa,L_{\rho,\mu}+\kappa}(V)$ to $\theta_{\rho,\mu,\kappa}$, with a good choice for ρ, μ and κ, we have

$$\theta(z_k) - \theta^* \to 0 \text{ in } O\left(\frac{1}{\epsilon}\ln\left(\frac{1}{\epsilon}\right)\right)$$

and

$$\|\nabla\theta_{\rho,\mu,\kappa}(z_k)\|_{V^*} \to 0 \text{ in } O\left(\frac{1}{\epsilon}\ln\left(\frac{1}{\epsilon}\right)\right).$$

Modified Dual Objective Function

$$\theta_{\rho,\mu,\kappa}(z) = \sigma_{\rho,T}(z) + \phi_\mu(z) + \frac{\kappa}{2}\|z\|_V^2$$

where

- $\sigma_{\rho,T}(z) = \sup_{x\in T}\{\langle x, z\rangle - \frac{\rho}{2}\|x\|_{V^*}^2\}$
- $\phi_\mu(z) = \sup_{u\in U}[-J(u) - \langle \mathcal{A}u, z\rangle - \frac{\mu}{2}\|u\|_U^2]$
- $\rho, \mu, \kappa > 0$.

This function is:

- Strongly convex with parameter κ
- Smooth with a Lipschitz-continuous gradient (constant $L_{\rho,\mu,\kappa} = L_{\rho,\mu} + \kappa$).

Why a double smoothing ? Summary

Method	Dual function	Dual conv.	Primal conv.
Subgradient	Convex but Non-Smooth	$O\left(\frac{1}{\epsilon^2}\right)$	$O\left(\frac{1}{\epsilon^2}\right)$
Simple Smoothing	Convex ∇ Lipschitz-cont	$O\left(\frac{1}{\epsilon}\right)$	$O\left(\frac{1}{\epsilon^2}\right)$
Double Smoothing	Strongly convex ∇ Lipschitz-cont.	$O\left(\frac{1}{\epsilon}\ln\left(\frac{1}{\epsilon}\right)\right)$	$O\left(\frac{1}{\epsilon}\ln\left(\frac{1}{\epsilon}\right)\right)$

Outline

1. Studied problem class.
2. Dual Approach.
3. Double Regularization.
4. Solving the dual problem.
5. Reconstruction of a primal solution.
6. Applications and Further Research.

Optimal Scheme for $S^{1,1}_{\kappa,L}(V)$

Let $g : V \to \mathbb{R}$ be

- strongly convex with parameter $\kappa > 0$
- smooth with a Lipschitz-continuous gradient with constant $L > 0$.

Algorithm

1. Initialization
   Choose $w_0 = z_0 \in V$.
2. Iteration $(k \geq 0)$
   Set

$$z_{k+1} = w_k - \frac{1}{L}\nabla g(w_k)$$

$$w_{k+1} = z_{k+1} + \frac{\sqrt{L} - \sqrt{\kappa}}{\sqrt{L} + \sqrt{\kappa}}(z_{k+1} - z_k).$$

Complexity

Let $\epsilon > 0$ and choose $\rho(\epsilon) = C_1\epsilon, \mu(\epsilon) = C_2\epsilon$ and $\kappa(\epsilon) = C_3\epsilon$, after

$$k(\epsilon) = O\left(\frac{1}{\epsilon}\ln\left(\frac{1}{\epsilon}\right)\right)$$

iterations, we have:

- $$\theta(z_{k(\epsilon)}) - \theta^* \leq \epsilon$$
- $$\left\|\nabla\theta_{\rho,\mu}(z_{k(\epsilon)})\right\|_{V^*} \leq \frac{2}{R}\epsilon.$$

Outline

1. Studied problem class.
2. Dual Approach.
3. Double Regularization.
4. Solving the dual problem.
5. Reconstruction of a primal solution.
6. Applications and Further Research.

A nearly feasible and optimal primal solution

Using the dual iterate $z_{k(\epsilon)}$, we can obtain a primal solution with the same order of accuracy.
Consider

$$u_{k(\epsilon)} = u_{\mu(\epsilon), z_{k(\epsilon)}}$$

the unique optimal solution of the problem defining $\phi_{\mu(\epsilon)}(z_{k(\epsilon)})$ that we can compute in closed-form.
This primal solution $u_{k(\epsilon)}$ is:

- In S by construction
- Nearly optimal for the primal problem:

$$\left| J(u_{k(\epsilon)}) - P^* \right| \leq 2(1 + 2\sqrt{3})\epsilon$$

- Nearly feasible for the linear constraint:

$$\operatorname{dist}(\mathcal{A}u_{k(\epsilon)}, T) \leq \|\nabla \theta_{\rho,\mu}(z_k(\epsilon))\|_{V^*} = \left\| \mathcal{A}u_{k(\epsilon)} - x_{\rho, z_k(\epsilon)} \right\|_{V^*} \leq \frac{2}{R}\epsilon.$$

Outline

1. Studied problem class.
2. Dual Approach.
3. Double Regularization.
4. Solving the dual problem.
5. Reconstruction of a primal solution.
6. Applications and Further Research.

Applications

The double smoothing scheme can be applied for solving:

- Separable Large-scale finite-dimensional problems with coupling constraint :

$$\min_{u=(u_1,\dots,u_N)} \sum_{i=1}^{N} J_i(u_i) : \quad \sum_{i=1}^{N} A_i u_i \in T \quad u_i \in U_i \quad \forall i = 1, \dots, N.$$

- Infinite-dimensional problems with coupling and pointwise constraint:

$$\min_{u \in L^2([0,T],\mathbb{R}^m)} \int_0^T F(t, u(t))dt$$

$$\int_0^T A(t)u(t)dt \in T$$

$$u(t) \in S(t) \text{ a.e. in } [0, T]$$

Applications (2)

- Optimal Control problems of the form:

$$P^* = \inf_{u \in L^2([0,T],\mathbb{R}^m)} \int_0^T G(t, u(t)) + \langle a(t), x(t) \rangle dt$$

$$\dot{x}(t) = A(t)x(t) + B(t)u(t), \quad x(0) = x_0$$

$$x(t_i) \in \overline{T}_i \quad i = 1, ..., N$$

$$u(t) \in S(t) \quad \text{a.e. in } [0, T].$$

Further research

- **Consequence if the subproblems defining $\sigma_{\rho,T}(z)$ and $\phi_\mu(z)$ can be solved only approximately using a FOM ?**

  Which accuracy do we need for solving the subproblems ?
  What is the total complexity (OUTER and INNER iterations) of the double smoothing approach ?

- **Comparison with other approaches for solving our problem class.**
  Augmented Lagrangian approach, Exact Penalty approach,...

Thanks for your attention !

UCL
Université catholique de Louvain