Conjugate Epi-Projection Algorithms for Non-Smooth Optimization and Related Issues

Evgeni Nurminski
Far Eastern Federal University, Vladivostok

Huawei RRI Operations Research and Mathematical Optimization Workshop
Moscow, September 27-29, 2020

Outline

- Non-smooth optimization
  - Motivations;
  - The high-speed conjugate epi-projection (CEP) algorithms;
  - CEP implementation;
- Projection;
  - Polytopes and polyhedrons;
  - Decomposition
  - Linear optimization
- Ongoing and planed work.

Motivations

- Decomposition: $obj(x) = subobj_1(x) + subobj_2(x) + \dots,$

$$subobj_i(x) \Leftarrow model_i(x, \dots), i = 1, 2 \dots;$$

- Reduction: $\min_{(x,y)\in Z} obj(x, y) = \min_x reduced.obj(x),$

$$reduced.obj(x) = \min_{y \in Z(x)} obj(x, y);$$

- Data compression, automatic classification;
- Request for robustness ($\min\max$ problems);
- Exact penalties, lagrangian relaxation;
- etc.

Conjugate subgradient algorithms

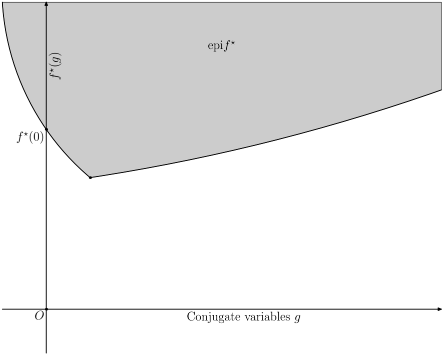
Descent direction is found as projection on $\mathrm{co}\{g^s, s = 1, 2, \dots\}$:

1. Wolfe, P.: A Method of Conjugate Subgradients for Minimizing Nondifferentiable Functions. Mathematical Programming Study, 3, 145-–173 (1975)
2. Li, Q.: Conjugate gradient type methods for the nondifferentiable convex minimization. Optimization Letters, 7(3), 533—545 (2013)

The same idea can be used for gradient methods for VI.

Convex analysis

The problem: $\min_x \; f(x), \quad f : E \to R_\infty,$ convex.



Epigraph of a conjugate function $f^\star(g) = \sup_x\{xg - f(x)\}$.

The basic idea:

$$f^\star(0) = -\min_x f(x) = -f_\star = \inf_{(0,\mu)\in \mathrm{epi}\, f^\star} \mu.$$

Conjugate Epi-Projection Algorithm

The algorithm consists of two basic operations:

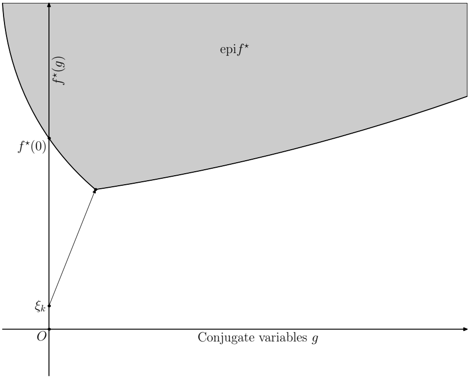
1. **Projection.**

$$\min_{(\xi,g)\in \mathrm{epi}\, f^\star} \{(\xi-\xi_k)^2+\|g\|^2\}.$$

2. **Support-Update.**
Compute support function $v_k = (\mathrm{epi}\, f^\star)_{z^k}$ and update the approximate solution with ξ_{k+1}

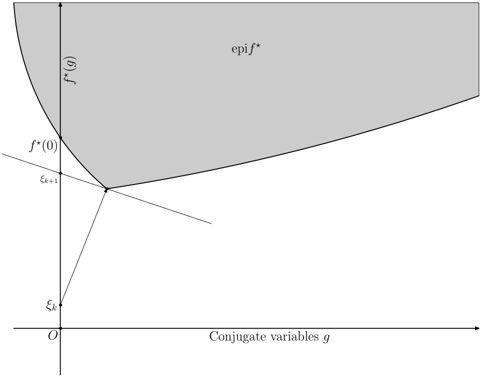
$$\xi_{k+1} = v_k/(f^\star(g_p^k)-\xi_k).$$

Project



Projection of $(\xi, 0)$ onto epi $f^\star$.

Support-Update



Compute support function of epi $f^\star$:

$$\sup_g \{x(z^k/\xi_k) - f^\star(g)\} = f(z^k/\xi_k)$$

Major convergence results

Proved:

- If $f(x)$ is just convex the convergence is superlinear:

$$f_{k+1} - f_\star \leq \lambda_k(f_k - f_\star),\ \lambda_k \to 0 \text{ when } k \to \infty$$

- If $f(x)$ is sup-quadratic the convergence is quadratic:

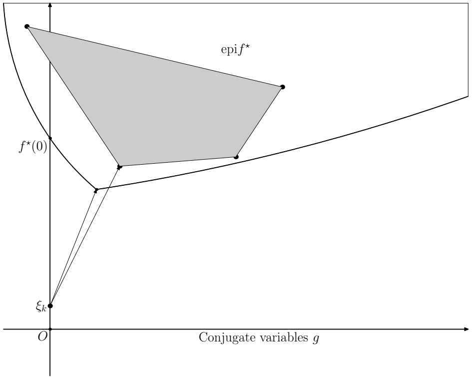
$$f_{k+1} - f_\star \leq \lambda(f_k - f_\star)^2, \quad \text{when } k \to \infty$$

  when $\lambda < f_0 - f_\star$ which garantees convergence.

- If $f(x)$ has *sharp minimum* then convergence is finite.

In all cases convergence is global, ie does not depend on initial point.

Implementable version



Approximate $\mathrm{epi}\, f^\star$ with $P_m = \mathrm{co}\{z^1, z^1, \dots, z^m\}$ and project.

Practicalities

The subproblem for projection polyhedron P_m can be solved by many off-the-shelf quadratic solvers, however our experience is that the specialized algorithms like

> *Nurminski, E.A. Convergence of the Suitable Affine Subspace Method for Finding the Least Distance to a Simplex. Computational Mathematics and Mathematical Physics, 45(11), 1915–1922 (2005)*

outperforms them.

One can download the PYTHON and/or OCTAVE versions of the code as `DOI: 10.13140/RG.2.2.21281.86882` from `ResearchGate`.

PTP algorithm

Data: $\hat{P} = \{\hat{p}^1, \hat{p}^2, \ldots, \hat{p}^N\}$
Result: $p^\star \in P = \mathrm{co}(\hat{P})$ with the minimal norm
Define initial $\bar{P} \subset \hat{P}$ and the least norm $\bar{p} \in \mathrm{lin}(\bar{P})$ such that $\bar{x} \in \mathrm{co}(\bar{P})$;
while *There is a chance to improve* $\bar{p}$ **do**

- Add some $\hat{p} \in \hat{P}$ which results in decrease of distance:

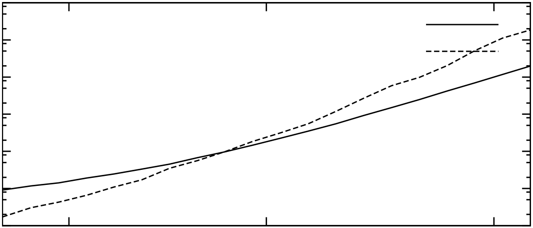
$$\min_{p \in \mathrm{lin}(\hat{p}, \bar{P})} \|p\| = \|p^s\| < \|\bar{p}\|$$

- Delete $\hat{p} \in \bar{P}$ with negative baricentrics.

end

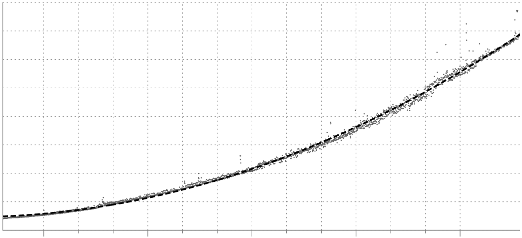
Run-Time Results

- QP – off-the-shelf general purpose quadratic programming subroutine.
- PTP – specialized polytope projection.



Run-time dependence on the rows-columns size of X.

PTP iterations complexity



PTP run-time dependence on the base size, fitted with the quadratic approximation $1.833\ 10^{-8}x^2 + 5.764\ 10^{-6}x + 0.0097$

Computational experience: CONDOR v CEP

Max-quadratic function:

$$f(x) = \max_{i=1,2}(x - a^i)A_i(x - a^i)$$

with $a^1 = (0, 0, 0), a^2 = (2, 3, 9)$ and diagonal matrices A_i:
$A_1 = \mathrm{diag}(9, 4, 1), A_2 = \mathrm{diag}(1, 4, 9)$.

CONDOR 1.06 (NEOS)	63	`0.4348696068`
CEP	27	`0.43673`

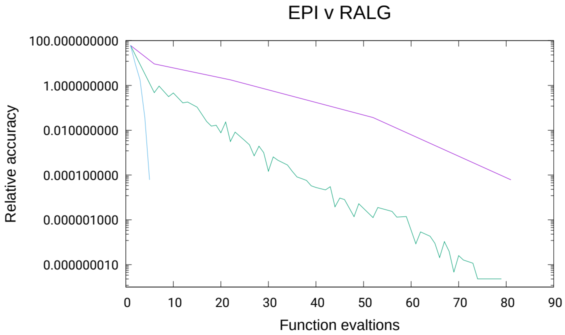
Computational experience: RALG v CEP

Test function:

$$f(x) = \max\{q_1(x), q_2(x)\}$$

where:

$$q_1(x) = x_1^2/25 + x_2^2/4 + x_3^2/49,$$
$$q_2(x) = (x_1 - 2)^2 + (x_2 - 3)^2/9 + (x_3 - 1)^2/25$$



Foundations

Orthogonal projection (the most common):

$$\min_{x \in X} \|x - a\|^2 = \|x^\star(a) - a\|^2 = \|\Pi_X(a) - a\|^2$$

where $\Pi_X(a) \in X$.

Good news:

a) $\Pi_X : E \to X$ — single-valued (follows from strong convexity).
b) Lipschitz continious with the Lipschitz constant $L_X \leq 1$: $\|\Pi_X(a) - \Pi_X(b)\| \leq L_X\|a - b\|$ for any a, b.

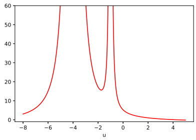
Not so good news:

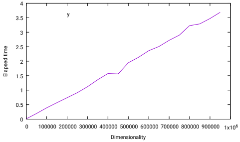
a) It is not so rare that $L_X = 1$ (nonexpansion) so forget about iteration algorithms.
b) Even if for X the constant $L_X < 1$ it may be VERY close to 1 so iteration algorithm may be VERY slow.

Trivial cases

- boxes, spheres, halfspaces, linear manyfolds — closed form solutions. Problems become nontrivial for huge dimensions, and/or degenerate cases but this is another story.
- ellipsoid – reducable to 1-dimensional polynom root finding problem with good bounds for the single positive real root. Smth like $n \log(\epsilon)$ complexity bound for ϵ-accuracy.

  Dual function for ellips projection $\psi(u) = \sum_{i=1}^{n} \frac{z_i^2}{a_i^2(1+u/a_i^2)^2} = 1$





About 1 mln variables — approx 3.5 sec.

Canonical simplex

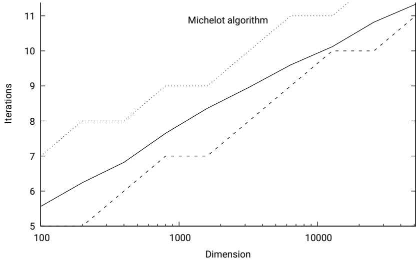
Projection problem with many applications $X = \Delta_E$

$$\min_{x \in \Delta_E} \|a - x\|^2.$$

The number of faces exponential in dimension n, the lowest algorithmic upper complexity bound is unknown. Algorithms with smth like $n \log(n)$ complexity:

- Michelot (C. Michelot, JOTA, 1986)
- Malozemov–Tamasyan, Comput. Math. and Math. Phys., 2016)
- and probably many others ...

Michelot algorithm




```
function [ x iter ] = michelot(z, rho)
x = z;
x += (rho - sum(x)) / rows(x);
iter = 0;
do
        bv  = (x > 0);  nbv = sum(bv);
        if !all(bv)
                x(!bv) = 0;
                x(bv) += ( (rho-sum(x(bv))) / nbv );
        endif
        iter++;
until all( x >= 0)
endfunction
```

Polytope projection

Problem: $\min_{x \in P} \|x\|^2$, where $P = \mathrm{co}\{\hat{p}^i, i \in I\} = \mathrm{co}\{\hat{P}\}$.

Rewrite as constrained QP ?

- $P = \{x : Qx \leq q\}$? Q may have an exponential number of rows !
- $\min \|x\|^2$ s.t. $x = \hat{P}s, s \in \Delta$. ? Essential increase in the number of unknowns. Semidefinite.
- Rewrite in baricentric coordinates ?

$$\min s\hat{P}^T\hat{P}s \text{ s.t. } s \in \Delta.$$

High chances of dense $\hat{P}^T\hat{P}$, not all $p^i p^j$ will actually be needed. May be semidefinite.

This motivated the development of a special algorithm not unlike the Active Set variety but with its own add-delete rules.

PTP algorithm

Data: $\hat{X} = \{\hat{x}^1, \hat{x}^2, \ldots, \hat{x}^N\}$
Result: $x^\star \in X$ with the minimal norm
Define initial $\bar{X} \subset \hat{X}$ and the least norm $\bar{x} \in \mathrm{lin}(\bar{X})$ such that $\bar{x} \in \mathrm{co}(\bar{X})$;
while *There is a chance to improve* $\bar{x}$ **do**

- Add some $\hat{x} \in \hat{X}$ which results in decrease of distance:

$$\min_{x \in Lin(\hat{x}, \bar{X})} \|x\| = \|x^s\| < \|\bar{x}\|$$

- Delete $\hat{x} \in \bar{X}$ with negative baricentric coordinate.

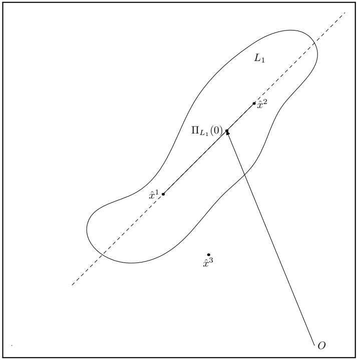
end

Nurminski E.A. *Convergence of the Suitable Affine Subspace Method …*: Comp. Math. Math. Phys., Vol. 45 No. 11, 2005, pp. 1915-1922.

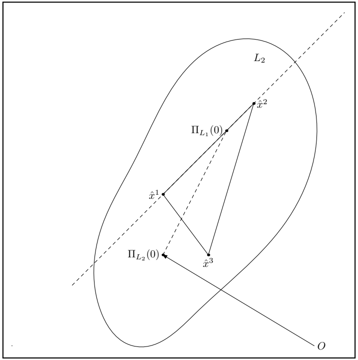
Python and Octave codes. `https://www.researchgate.net`, my page.

Exercise in Geometry

Start from a *suitable* basis



Halfway to the next *suitable* basis

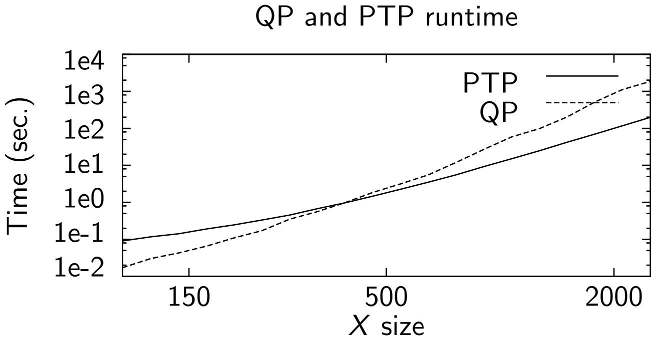


A suitable basis for $X = \{\hat{x}^1, \hat{x}^2, \dots\}$ is such subset $Y \subset X$ that

$$\min_{x \in \mathrm{Lin}(Y)} \|x\| = \min_{x \in \mathrm{co}\{Y)\}} \|x\|$$

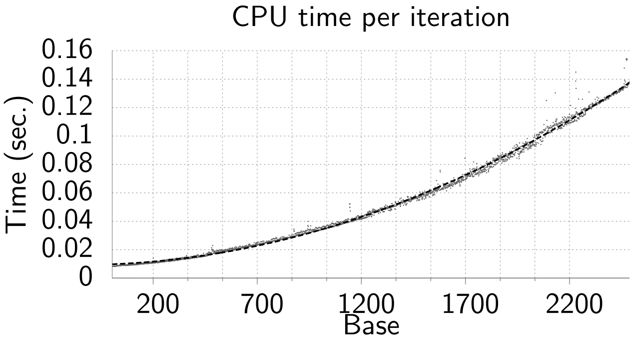
Run-Time Results

- QP – off-the-shelf general purpose quadratic programming subroutine.
- PTP – specialized polytope projection.



Run-time dependence on the rows-columns size of X.

PTP iterations complexity



PTP run-time dependence on the base size, fitted with the quadratic approximation $1.833\ 10^{-8}x^2 + 5.764\ 10^{-6}x + 0.0097$

Consider LO-problem:

$$\min_{Ax \leq b} cx = cx^{\star}.$$

Seems everybody knew but nobody cared to proof that

$$x^{\star} = \Pi_X(x^0 - \theta c)$$

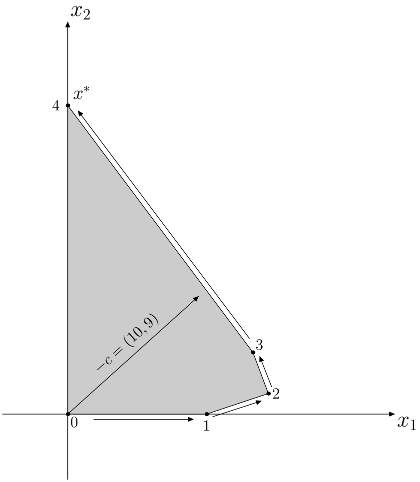
for arbitrary x^0 and large enough $\theta > 0$.

Lemma. Let $x^{\star}, u^{\star}$ are unique primal-dual solutions of the primal-dual LO formulations of the problem above, which satisfy strict complementarity conditions

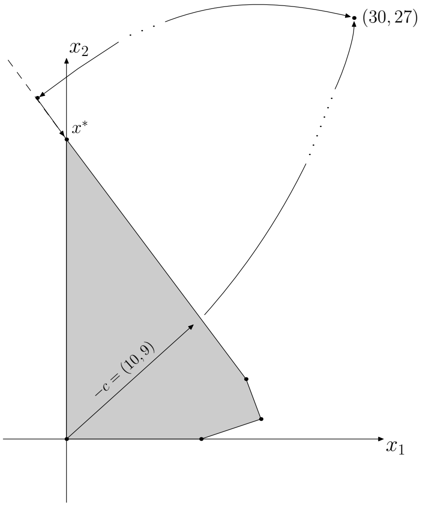
$$u^{\star}(Ax^{\star} - b) = 0; \;\; u^{\star} > Ax^{\star} - b$$

and $K_X^{\circ}(x^{\star})$ is a polar cone for the feasible set X at the optimal point $x^{\star}$. Then $-c \in \text{int}(K_X^{\circ}(x^{\star}))$.

Linear optimization



Simplex



Single-projection procedure, $\theta = 3$

Linear optimization: polyhedrons

Traditionly

$$\text{LO fesible set} = \{x : Ax \leq b\},$$

conversion to polytopes problematic if possible at all.
However it can be reduced to the cone projection:

$$\min_{z \in \mathrm{Co}\{\bar{A}_i, i=1,2,\ldots,m\}} \|z-a\|^2$$

where z is $(n+1)$–dimensional variable, $\bar{A}_i$ – *almost* i-th row of A. See

1. Nurminski E.A., Projection onto Polyhedra in Outer Representation Computational Mathematics and Mathematical Physics, 2008, Vol. 48, No. 3, pp. 367-375.
2. Evgeni Nurminski, Replacing projection on finitely generated convex cones with projection on bounded polytopes, arXiv:2010.12365 [math.OC], 2020.

Polytope Decomposition

Data: $A = \{a^i, i = 1, 2, \dots, m\}$, and $A_k \subset A, k = 1, K$ such that $A = \cup_{k=1,2,\dots,K} A_k$.

Result: $x^\star \in \mathrm{co}\{A\}$ such that $\|x^\star\| = \min_{x \in \mathrm{co}\{A\}} \|x\|$

while *There is a chance to improve* $x^\star$ **do**

- Decompose:

$$\min_{x \in Conv\{A_k, x^\star\}} \|x\|^2 = \|x^k\|^2, \; k = 1, 2, \dots, K.$$

- Coordinate:

$$\min_{x \in Conv\{x^k, k=1,2,\dots,K\}} \|x\|^2 = \|x^\star\|^2.$$

end

Planned developments

- Nonsmooth optimization and variational inequalities:
  - Implementable CEP;
  - Skew and multiple cuts;
  - Low-dimensional CO.
- Linear optimization:
  - Fine-grade, dynamic and nested decomposition;
  - Large-scale production applications;
  - Parallel computations.