Practical Projection with Applications

Evgeni A. Nurminski
nurminskiy.ea@dvfu.ru

Far Eastern Federal University, Vladivostok

Variational Analysis and Optimisation Webinar
Mathematics of Computation and Optimisation
Special Interest Group of the AustMS

Online August 5, 2020

Outline

- Basics notations and properties
- Simple sets
- Polytopes
    - PTP Algorithm
    - Numerical performance
- Polyhedrons
    - Reduction to cone projection
    - Numerical performance
    - Scaled truncated cones approach
- Applications — promises and problems
    - Linear Optimization
    - Decomposition
    - Nondifferentiable optimization

Standard notations

- $\mathbb{R}$ — the real axis with elements $\alpha, \beta, \ldots, \omega$. $\mathbb{R}_+$ is the nonnegative part of $\mathbb{R}$.
- $E_n = \overbrace{\mathbb{R} \times \mathbb{R} \times \ldots \mathbb{R}}^{n}, \quad E_n^+ = \overbrace{\mathbb{R}_+ \times \mathbb{R}_+ \times \ldots \mathbb{R}_+}^{n}$. Dimension n may be omitted if irrelevant.
- $\mathbf{0}$ and $\mathbf{1}$ are null and unit $(= (1, 1, \ldots, 1))$ elements of E.
- ab — the inner product of $a, b \in E$. Orthonorm $\|a\| = \sqrt{aa}$, $\|x\|_\infty = \max_{i=1,2,\ldots,n} |x_i|$.
- $\Delta_E = \{x : \mathbf{1}x = 1, x \in E^+\}$ — standart symplex in E. Can also be denoted as $\Delta_n \subset E_n^+$ if it matters.

Convexity

Here A is a finite set of ponts $a^1, a^2, \dots, a^m$ in E.

- $\mathrm{co}\, A = \{x = \sum_{i=1}^m \alpha_i a^i\}$ with $\alpha \in \Delta_m$ is a convex hull of A. It can also be called polytope;
- $\mathrm{Co}\, A = \{x = \sum_{i=1}^m \alpha_i a^i\}$ with $\alpha \in E_m^+$ is a conical hull of A. It can also be called convex polyhedral cone;
- A defines also the linear operator $A : E_m \to E_n$ such that $Az = x = \sum_{i=1}^m z_i a^i$. $AZ = \{Az, z \in Z$;
- $(X)_x = \min_{z \in X} xz$ is the support function of X. Finite for closed convex bounded sets (default);
- $\mathrm{epi}\, f = \{(x, \mu) : \mu \geq f(x)\} \subset E \times \mathbb{R}$ — the epigraph of function f;
- Fenchel-Moreau conjugate function: $f^\star(g) = \sup_x\{xg - f(x)\} = (\mathrm{epi}\, f)_{(g,-1)}$.

Polytopes and polyhedrons

Let P is the default set.

- If $P = \mathrm{co}\{A_P\}$ for some finite set A_P then it is called the *inner* representation of P.
- If $P = \{x : A_P x \leq b_P\}$ for some linear operator A_P and $b_P \in E_m$ then it is called the *outer* representation.

These representations are equivalent, however they may have very different complexity: any of these may have exponential (wrt dimension) complexity with the polynomial counterpart.

It advocates different algorithmic approaches for solving computational problems in these two representations.

Projection problem

Orthogonal projection (most common):

$$\min_{x \in X} \|x - a\|^2 = \|\Pi_X(a) - a\|^2 = \min_{x \in X - a} \|x\|^2 = \|\Pi_{X-a}(\mathbf{0})\|^2$$

where $\Pi_X(a) \in X$ and where $\Pi_{X-a}(\mathbf{0}) \in X - a$.

Good news:

a) $\Pi_X : E \to X$ — single-valued (follows from strong convexity).

a) 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:

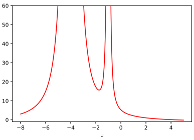
a) It is not so rare that $L_X = 1$ (nonexpansion) so forget about iteration algorithms.

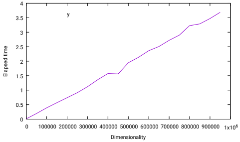
b) Even if for some X constant $L_X < 1$ it may be VERY close to 1 so iteration algorithm may be VERY slow.

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

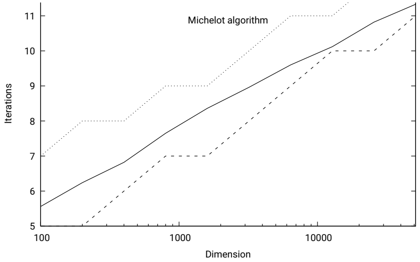
Projection problem with many applications $X = \Delta_E$

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

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 ?**

$$\min \|x\|^2 \text{ s.t. } x = \hat{P}s, s \in \Delta.$$

  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: $\hat{X} = \{\hat{x}^1, \hat{x}^2, \dots, \hat{x}^N\}$
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.

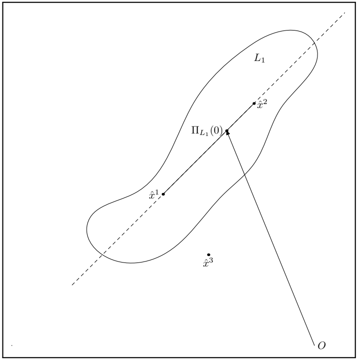
end

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

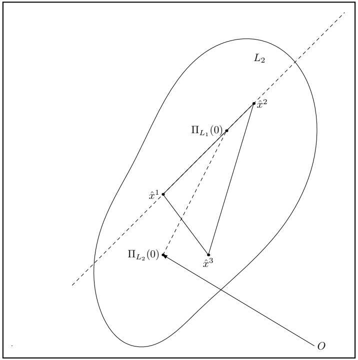
Python and Octave codes. `https://www.researchgate.net`, my page.

Exercise in Geometry

Start from a *suitable* basis



Halfway to the next *suitable* basis

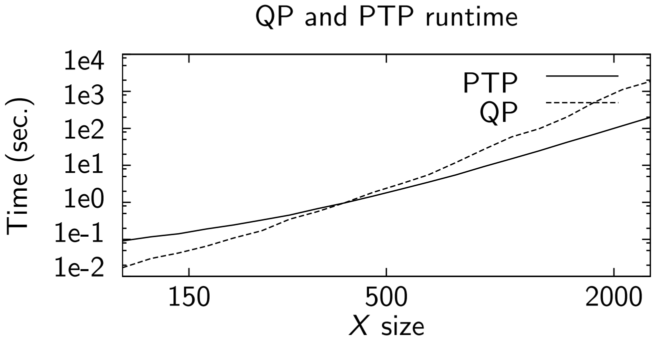


A suitable basis for $X = \{\hat{x}^1, \hat{x}^2, \dots\}$ is such subset $Y \subset X$ that

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

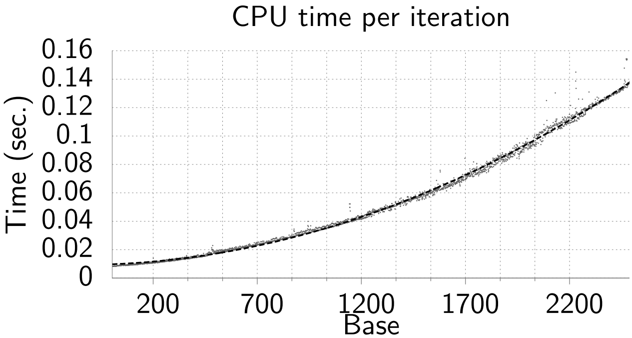
Run-Time Results

- 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 $1.833\ 10^{-8}x^2 + 5.764\ 10^{-6}x + 0.0097$

Systems of linear inequalities

What about $\min_{x \in X} \|x\|^2$ when

$$X = \{x : Ax \leq b\} = \{x : a^i x \leq \beta_i,\ i = 1, 2, \ldots, m\}?$$

Problems:

- direct transformation into polytope

  $$P_X = \mathrm{co}\{\hat{x}^k, k = 1, 2, \ldots, K\}$$

  is impractical because of exponentialy large K.
- Something like column generation technique with solving LP problems of the type

  $$\min_{x \in X} px = \min_{Ax \leq b} px$$

  also does not look very promissing.

But convex analysis comes to the resque

A few simple transformations

$$Ax \leq b \quad \leftrightarrow \quad \begin{array}{l} \bar{A}\bar{x} \leq 0 \\ a\bar{x} = 1 \end{array}$$

where $\bar{A} = |A, \ -b|, \bar{x} = (x, \xi)$ and $a = (0, 0, \ldots, 0, 1)$.
Redenote for simplicity:

$$K = \{x : Ax \leq 0\}, \qquad H = \{x : xa = 1\}$$

where $a, x \in E$.

Continue on the next 10 slides ...

Using standard duality we can suggest the following sequence of transformations:

$$\min_{x \in K_A \cap H} \tfrac{1}{2}\|x\|^2 = \min_{\substack{x^1 \in K_A, x^2 \in H, \\ x^1 = x^2}} \tfrac{1}{2}\|x^1\|^2 =$$

$$\max_{\theta,u}\{-\theta + \min_{x^1,x^2 \in K}\{\tfrac{1}{2}\|x^1\|^2 + u(x^1 - x^2) + \theta a x^1\}\} =$$

$$\max_{\theta}\{-\theta + \max_u\{\min_x\{\tfrac{1}{2}\|x\|^2 + (u + \theta a)x\} + \min_{x \in K}\{-ux\}\}$$

Of course

$$\min_x\{\frac{1}{2}\|x\|^2 + (u + \theta a)x\} = -\frac{1}{2}\|u + \theta a\|^2$$

and $\min_{x \in K}\{-ux\}$ is the indicator function of K^+ (upto taking into account $-u$), so we arrive to the next slide ..

Final result

Taking out all intermediate computation we see the final result as

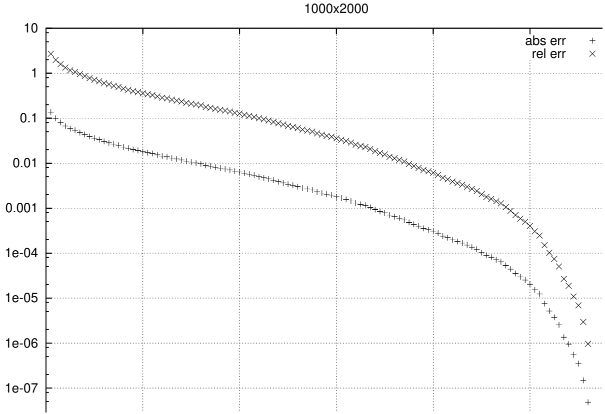
$$\min_{x \in K_A \cap H} \frac{1}{2}\|x\|^2 = \max_{\theta, u \in K^+}\{-\theta - \frac{1}{2}\|u - \theta a\|^2\} =$$
$$- \min_\theta\{\theta + \min_{u \in K^+} \|u - \theta a\|^2\}$$

The essential part of it is just projection of the point θa on the cone K^+:

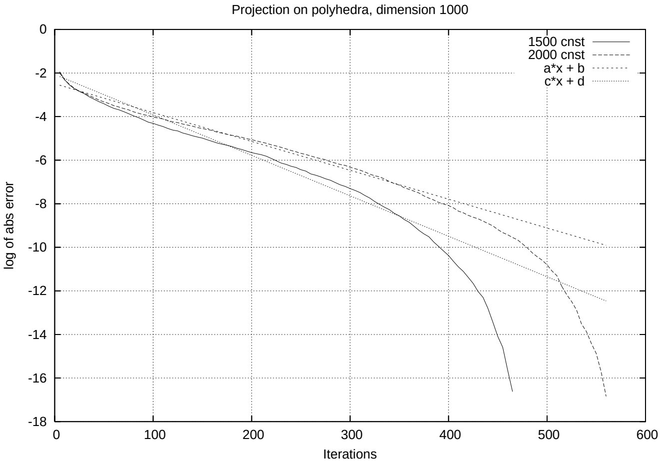
$$\phi(\theta) = \min_{u \in K^+} \|u - \theta a\|^2 = \phi(1)\theta^2.$$

It means that to project on a polyheadron we have to project the point $a = (0, 0, \ldots, 0, 1)$ on the cone gererated by rows of the system of inequalities (in fact in conjugate space)

Rate of convergence for polyheadron 1000x2000



Rate of convergence for polyheadron projection

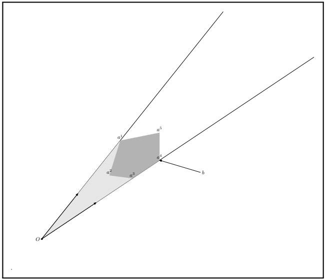


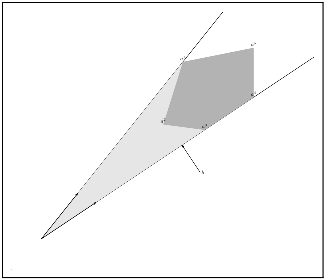
Projection on random polyheadrons

n	m	min	max	ave	std	ntest	fail
1000	1500	470	490	476.7143	7.0170	7	0
1000	1600	478	521	500.3000	14.5911	10	0
1000	1700	502	541	523.1000	11.3964	10	0
1000	1800	516	558	537.3000	14.8702	10	0
1000	1900	549	581	567.5556	11.7698	9	0
1000	2000	568	596	581.2857	10.7659	7	0
1500	1800	535	567	554.0000	12.9228	5	0
1500	1900	567	595	576.7500	12.7639	4	0
1500	2000	574	616	596.7500	17.6517	4	0
1500	2100	595	631	617.6667	19.7315	3	0

n — vars, *m* — ineqs, *min*, *max* — min,max iters, *ave* — mean iters, *std* — std dev, *ntest* — succs, *fail* — fails.

Truncated cones



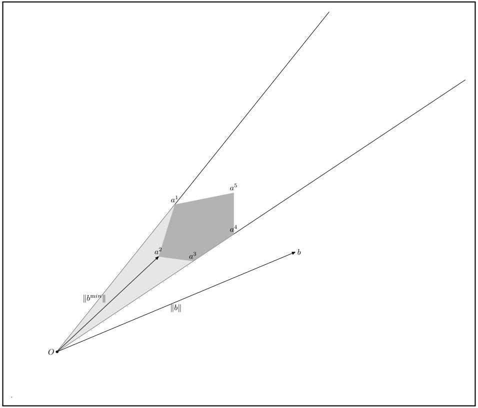


It is almost obvious that for scaling factor γ large enough

$$\min_{x \in \mathrm{Co}\{A\}} \|x - b\|^2 = \min_{x \in \gamma\,\mathrm{co}\{,A\}} \|x - b\|^2$$

. **Q:** How big must be γ ?

Truncated cones



Thm: If $bb^{\mathrm{m}in} < \|b^{\mathrm{m}in}\|^2$ then $\Pi_{A_c}(b) = \Pi_{\mathrm{Co}\{A\}}(b)$.
It implies that $\gamma > \|b\|/\|b^{\mathrm{m}in}\|$ will suffice.

Scaled truncated cone algorithm

Data: The set $A = \{a^i, i = 1, 2, \dots\}$ of generators of a cone $K(A)$, the vector b to be projected on the cone $K(A)$.

Result: The solution b^K of projection problem: $b \to \mathrm{Co}\{A\}$.

Phase 1. Compute a suitable value for the scaling parameter ρ by solving the auxiliary polytope projection problem

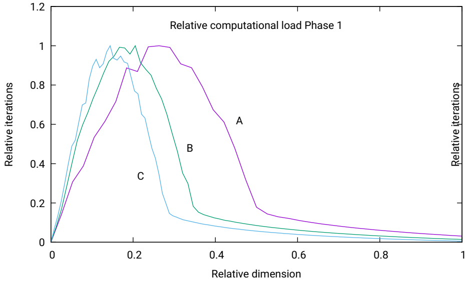
$$\min_{z \in \mathrm{co}\{A\}} \|z\|^2 = \rho^2_{\mathrm{m}in}$$

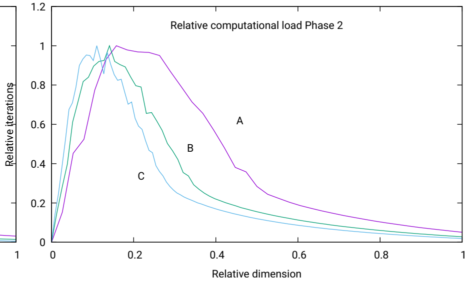
and set $\rho > \|b\| / \rho_{\mathrm{m}in}$.

Phase 2: Solve the projection problem

$$\min_{z \in \mathrm{co}\{\mathbf{0}, \rho A_c\}} \|z - b\|^2 = \|b^K - b\|^2.$$

Numerical experiments with STAC





Solution of the problem phase-1 Solution of the problem phase-2.

Computational complexity for solutions of problems phase-1, phase-2 for three different values of the size of the data set: A – 10^6, B – $2 \cdot 10^6$ and C – $3 \cdot 10^6$ dual precision elements.

Consider LO-problem:

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

Seems everybody knew but nobody cared to proof that

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

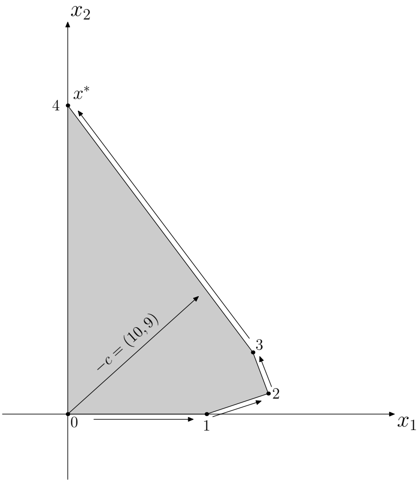
for arbitrary x^0 and large enough $\theta > 0$.

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

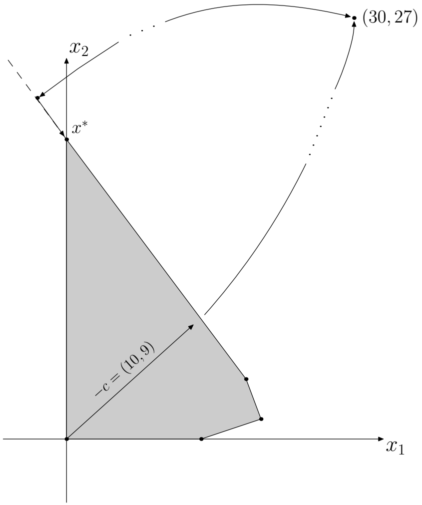
$$u^\star(Ax^\star - b) = 0;\ \ u^\star > Ax^\star - b$$

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

Linear optimization



Simplex



Single-projection procedure, $\theta = 3$

Polytope Decomposition

Let $A = \{a^i, i = 1, 2, \ldots, m\}$ and consider $\min_{x \in X} \|x\|^2$ where

$$X = \mathrm{co}\{A\} = \mathrm{co}\{A^c_k, k = 1, 2, \ldots, K\},$$

and $A^c_k = \mathrm{co}\{A_k\}$, and $A_k \subset A, k = 1, K$ is a covering of A.
Algorithm for $\min_{x \in X} \|x\|^2$:
start with $z^0 \in A, m = 0$
Loop:
Decomposition: solve $\min_{w \in co\{A^c_k, z^m\}} \|w\|^2 = \|w^k\|^2, k = 1, 2, \ldots, K$
Coordination: solve $\min_{z \in \mathrm{co}\{w^k, k=1,2,\ldots,K} \|z\|^2 = \|z^{m+1}\|^2$
Work MUCH better if we adapt covering !

Conjugate subgradient algorithms

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

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.

Conjugate Epi-Projection Algorithm

The basic idea:

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

(recall that $f^\star(g) = \sup_x \{xg - f(x)\}$) and use projection onto the epigraph $\mathrm{epi}\, f^\star$ for computing $f^\star(0)$.
As the result the algorithm consists of two basic operations:

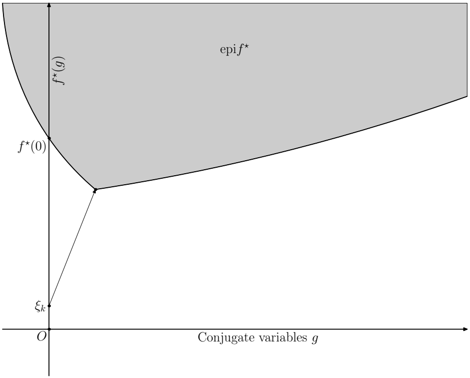
1. **Projection.**

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

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

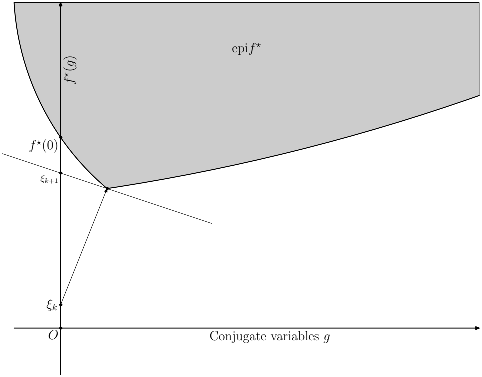
$$\xi_{k+1} = v_k/(f^\star(g_p^k) - \xi_k).$$

Project



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

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.

Related references

These are references which relate directly to the talk.

1. Nurminski, E.A.: A Conceptual Conjugate Epi-Projection Algorithm of Convex Optimization: Superlinear, Quadratic and Finite Convergence. Optim Lett(2019) 13:23-34 ISSN 1862-4472 DOI 10.1007/s11590-018-1269-3
2. Nurminski, E. A.: (2016). Single-projection procedure for linear optimization. Journal of Global Optimization, 66(1), 95–110. DOI:10.1007/s10898-015-0337-9 ISSN: 0925-5001 (Print) 1573-2916 (Online)
3. Vorontsova, E.A.; Nurminski E.A. Synthesis of cutting and separating planes in a nonsmooth optimization method Cybernetics and Systems Analysis, Vol. 51, No . 4, July, 2015, 619-631
4. Nurminski E.A. Replacing projection on finitely generated convex cones with projection on bounded polytopes. ResearchGate preprint, July 2020, DOI: 10.13140/RG.2.2.35735.19364.