ADGO 2013
Workshop on Algorithms and Dynamics for Games and Optimization
Playa Blanca, Tongoy, Chile — October 14th-18th, 2013

Open Problems

1. Samir Adly, *Pareto eigenvalue complementarity problem*
2. Jean-Bernard Baillon, *On the alternating projection method*
3. Luis Briceño, *Partial inverse and duality applied to monotone inclusions*
4. Patrick Combettes, *How to really split $m > 2$ operators?*
5. Roberto Cominetti, *Entropic convexity of expected utilities*
6. José Correa, *Adaptive shortest paths with probabilistic link failures*
7. Christoph Dürr, *Stealing strategies*
8. Jannik Matouschke, *The length-bounded VPN design problem*
9. Rida Laraki & P. Mertikopoulos, *Inertial gradient flows in convex optimization*
10. Alantha Newman, *Rank aggregation with partitioning*
11. Ivan Rapaport, *The connectivity problem in the* number-in-hand *computation model*
12. Sylvain Sorin, *Delegation equilibrium in atomic splitting congestion games*
13. Nguyen Kim Thang, *Lagrangian duality in online scheduling*
14. Jorge Vera, *Consistency of decisions at different time horizons*

Pareto eigenvalue complementarity problems

Samir Adly - Université de Limoges

A scalar $\lambda > 0$ is called a Pareto eigenvalue of a matrix $A \in \mathbb{R}^{n\times n}$ if there is $x \in \mathbb{R}^n \setminus \{0\}$ such that

$$0 \leq x \perp (Ax - \lambda x) \geq 0.$$

Pareto eigenvalues appear for instance in the stability analysis of finite dimensional elastic structures with frictional contacts. An open problem is to determine the maximum number of such eigenvalues. More precisely, denoting $\sigma_P(A)$ the set of Pareto eigenvalues of the matrix A, we want to determine

$$\pi_n = \max_{A\in\mathbb{R}^{n\times n}} card(\sigma_P(A)).$$

The best currently known bounds are

$$3(2^{n-1} - 1) \leq \pi_n \leq n2^{n-1} - (n - 1).$$

In particular $\pi_1 = 1$, $\pi_2 = 3$ and $\pi_3 = 9$ or 10. Note also that $\pi_{20} \geq 1\,572\,861$. The following matrices of order 3,4,5 have exactly 9, 23 and 57 Pareto eigenvalues respectively

$$A_3 = \begin{pmatrix} 5 & -8 & 2 \\ -4 & 9 & 1 \\ -6 & -1 & 13 \end{pmatrix} \qquad A_4 = \begin{pmatrix} 132 & -106 & 18 & 81 \\ -92 & 74 & 24 & 101 \\ -2 & -44 & 195 & 7 \\ -21 & -38 & 0 & 230 \end{pmatrix}$$

$$A_5 = \begin{pmatrix} 788 & -780 & -256 & 156 & 191 \\ -548 & 862 & -190 & 112 & 143 \\ -456 & -548 & 1308 & 110 & 119 \\ -292 & -374 & -14 & 1402 & 28 \\ -304 & -402 & -66 & 38 & 1522 \end{pmatrix}$$

Q1. Find a matrix $A \in \mathbb{R}^{3\times3}$ with 10 Pareto eigenvalues.

Q2. How to improve the lower bound?

Q3. Find the asymptotic growth order of π_n as n goes to infinity.

On the alternating projection method

Jean-Bernard Baillon

Problem 1. The solution of $\frac{du}{dt} = Mu$ is $u(t) = e^{tM}u(0)$. For $M = A + B$ in general $e^{t(A+B)} \neq e^{tA}e^{tB}$ while the Trotter-Lie formula gives

$$e^{t(A+B)} = \lim_n (e^{\frac{t}{n}A}e^{\frac{t}{n}B})^n.$$

When $A = \Delta$ and $B = V$ one can solve $\frac{du}{dt} + \Delta u = 0$ and $\frac{du}{dt} + V(u) = 0$, and then consider an alternating solution method. Note that $\Delta u + V(u)$ might not be well defined even if $V(u) = vu \ldots$ *open problem.*

Problem 2. On Hilbert space, consider a finite family of orthogonal projectors P_j onto F_j. We know that $(P_m \cdots P_1)^n x$ converges strongly to the projection of x onto $\cap_j F_j$. Can we accelerate this convergence? Amemiya & Ando (1964) looked at sequences $P_{\varphi(n)} \cdots P_{\varphi(1)}x$ with $\varphi(n) \in \{1, \ldots, m\}$, and gave conditions for weak convergence. Since then, many have tried to establish strong convergence: Bruck, Dye, Reich, jbb, Lin, PLL,... This would follow from Bruck's Conjecture: if C_m is the Halperin constant, then

$$\|P_{\varphi(n)} \cdots P_{\varphi(1)}x - x\|^2 \leq C_m(\|x\|^2 - \|P_{\varphi(n)} \cdots P_{\varphi(1)}x\|^2).$$

For $m = 2$ the best constant is 2. For $m = 3$ we do not know. Numerical tests using semi-definite programming have been used to estimate the best constant. Paszkiewicz (ArXiv, 2012) gave a quasi-definitive answer for $m = 5$, using simple tools. The cases $m = 3, 4$ are open. Also open is the case of self-adjoint operators and the characterization of strong convergence.

Problem 3. Alternate projections onto two convex sets C_1 and C_2 generate a sequence that converges to a pair attaining the minimum distance $d(C_1, C_2)$. For 3 or more sets, cyclic projections still converge but no variational characterization exists for the limit cycle (jbb-Combettes-Cominetti, JFA 2012).

Which is the variational formulation for $A+B$? In particular for $-\Delta u+fu$? In dimension 1, $-u''+fu$ with $f \geq 0$, $f \in L^1_{loc}$ but $f \notin L^2_{loc}$, the domain of the sum operator is 0. How it can be enlarged? For φ, ψ l.s.c. convex functions

$$\partial\varphi \oplus \partial\psi = \partial(\varphi + \psi).$$

What happens when $f \not\geq 0$? What about if A, B are no longer sub-differentials and/or nonlinear? What can be said for their difference $A - B$?

Problem 4. Let R_n denote the inverse of the lower triangular matrix $M_n = \left(\lfloor \frac{i}{j} \rfloor\right)_{1 \leq i,j \leq n}$, *e.g.*

$$M_3 = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 1 & 1 \end{bmatrix}$$

Can you prove that $|\sum R_{ij}| \leq \text{const}\sqrt{n}$? What about $|\sum R_{ij}| \leq k_\epsilon\, n^{\frac{1}{2}+\epsilon}$?

REMARK: If $\lfloor \frac{i}{j} \rfloor$ is replaced by $(\frac{i}{j})_+$ where $x_+ = 0$ if $x < 1$ and $x_+ = x$ otherwise, we have $|\sum R_{ij}| \sim \ln n$.

PRIZE: US$1.000.000. And maybe a Fields Medal?

ADGO 2013
Playa Blanca, Chile
14 – 18 October, 2013

PARTIAL INVERSE AND DUALITY APPLIED TO MONOTONE INCLUSIONS

Luis M. Briceño Arias[1]

[1]Universidad Técnica Federico Santa María

In 1983 Spingarn introduced the partial inverse of a maximally monotone operator $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ with respect to a closed vectorial subspace V of the real Hilbert space $\mathcal{H}$ by

$$\begin{cases} A_V : \mathcal{H} \to 2^{\mathcal{H}} \\ u \in A_V x \quad \Leftrightarrow \quad P_V u + P_{V^\perp} x \in A(P_V x + P_{V^\perp} u). \end{cases} \tag{1}$$

Note that $A_{\mathcal{H}} = A$ and $A_{\{0\}} = A^{-1}$. Spingarn talked about some relations between partial inverse and duality but not in a precise way. In classical monotone operator theory, the dual of the inclusion

$$\text{find } x \in \mathcal{H} \quad \text{such that} \quad 0 \in Ax + Bx, \tag{2}$$

where $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ are maximally monotone, is

$$\text{find } u \in \mathcal{H} \quad \text{such that} \quad 0 \in A^{-1}u - B^{-1}(-u). \tag{3}$$

It is not difficult to prove that another equivalent formulation using partial inverse with respect to V gives raise to the following inclusions in duality

$$\begin{cases} \text{find } v \in \mathcal{H} \quad \text{such that} \quad 0 \in A_V v + R_V \circ B_V(R_V v) \\ \text{find } y \in \mathcal{H} \quad \text{such that} \quad 0 \in B_{V^\perp} y - R_V \circ A_{V^\perp}(-R_V y), \end{cases} \tag{4}$$

where $R_V = 2P_V - \mathrm{Id}$ is the reflection operator with respect to V and y and v are related to x and u via

$$\begin{cases} v = P_V x + P_{V^\perp} u \\ y = P_V u + P_{V^\perp} x. \end{cases} \tag{5}$$

Note that if $V = \{0\}$ we obtain (2) and (3). Currently I am interested in the following questions.

Problem 1 It is well known that, under qualification conditions, when $A = \partial f$ and $B = \partial g$ for some convex lsc proper functions $f\colon \mathcal{H} \to \,]-\infty, +\infty]$ and $g\colon \mathcal{H} \to \,]-\infty, +\infty]$, (2) and (3) reduce

to

$$\begin{cases} \underset{x\in\mathcal{H}}{\text{minimize}} & f(x)+g(x) \\ \underset{u\in\mathcal{H}}{\text{minimize}} & f^*(u)+g^*(-u). \end{cases} \tag{6}$$

What are the primal dual optimization problems associated to (4) when $A = \partial f$ and $B = \partial g$? What are the duality objects that appear in this optimization setting?

Problem 2 In which instances to solve (4) could be better than solving (2) and (3)?

Problem 3 Suppose that $\mathcal{H} = U \oplus V \oplus W$ and consider the inclusion

$$\text{find } x \in \mathcal{H} \quad \text{such that} \quad 0 \in Ax + Bx + Cx, \tag{7}$$

where $A\colon \mathcal{H} \to 2^{\mathcal{H}}$, $B\colon \mathcal{H} \to 2^{\mathcal{H}}$, and $C\colon \mathcal{H} \to 2^{\mathcal{H}}$ are maximally monotone. Is there a way to write an equivalent formulation by using A_U, B_V, and C_W separately? What kind of algorithms for solving such system (if exist) will appear?

References

J. E. Spingarn, Partial inverse of a monotone operator, *Appl. Math. Optim.*, **10** (1983) 247–265.

Can One Genuinely Split $m > 2$ Monotone Operators?

P. L. Combettes
Laboratoire Jacques-Louis Lions, Faculté de Mathématiques
Université Pierre et Marie Curie – Paris 6, 75005 Paris, France
Playa Blanca – 14 Octubre 2013

Throughout $\mathcal{H}$ is a real Hilbert space and $\operatorname{zer} C = \big\{x \in \mathcal{H} \mid 0 \in Cx\big\}$ is the set of zeros of a set-valued operator acting on $\mathcal{H}$. Many problems in nonlinear hilbertian analysis can be reduced to

find $x \in \operatorname{zer} C$, where $C\colon \mathcal{H} \to 2^{\mathcal{H}}$ is maximally monotone.

This inclusion can be solved by the proximal point algorithm (the resolvent of C is $J_C = (\operatorname{Id} + C)^{-1}$)

$$x_{n+1} = J_{\gamma_n C} x_n, \tag{1}$$

where $(\gamma_n)_{n\in\mathbb{N}}$ lies in $]0, +\infty[$ and $\sum_{n\in\mathbb{N}} \gamma_n^2 = +\infty$ [3]. Unfortunately, in most situations, (1) is not implementable because the resolvents of C are too hard to compute. In *splitting methods,* we decompose C in terms of operators which are simpler (i.e., they can be used explicitly or have easily computable resolvents), and we devise an algorithm which employs these operators individually.

Consider the basic inclusion with two maximally operators $0 \in Ax + Bx$. There exist only 3 basic splitting methods to solve this inclusion [2]:

- **Douglas-Rachford algorithm:** $\gamma \in\ ]0, +\infty[$.
  - $\operatorname{zer}(A+B) = J_{\gamma B}\Big(\operatorname{Fix}\Big(\frac{1}{2}\big((2J_{\gamma A} - \operatorname{Id}) \circ (2J_{\gamma B} - \operatorname{Id}) + \operatorname{Id}\big)\Big)\Big)$.
  - Iterate

$$\left\lfloor \begin{array}{lll} x_n & = J_{\gamma B} y_n & \text{(backward step)} \\ y_{n+1} & = J_{\gamma A}\big(2x_n - y_n\big) + y_n - x_n & \text{(backward step)} \end{array} \right.$$

    Then $y_n \rightharpoonup y$, $z = J_{\gamma B} y \in \operatorname{zer}(A+B)$ [6], and $x_n \rightharpoonup z \in \operatorname{zer}(A+B)$ [2, 8].

- **Forward-Backward algorithm:** $\gamma \in\ ]0, +\infty[$.
  - $B\colon \mathcal{H} \to \mathcal{H}$ is β-cocoercive: $\langle x - y \mid Bx - By\rangle \geqslant \beta\|Bx - By\|^2$; $\gamma \in\ ]0, 2\beta[$.
  - $\operatorname{zer}(A+B) = \operatorname{Fix}\Big(J_{\gamma A}\big(\operatorname{Id} - \gamma B\big)\Big)$.
  - Iterate

$$\left\lfloor \begin{array}{lll} y_n & = x_n - \gamma B x_n & \text{(forward step)} \\ x_{n+1} & = J_{\gamma A} y_n & \text{(backward step)} \end{array} \right.$$

    Then $x_n \rightharpoonup z \in \operatorname{zer}(A+B)$ [7].

- **Forward-Backward-Forward algorithm:** $\gamma \in\ ]0, +\infty[$.
  - $\operatorname{zer}(A+B) = \operatorname{Fix}\Big(J_{\gamma A}\big(\operatorname{Id} - \gamma B\big)\Big)$.
  - $B\colon \mathcal{H} \to \mathcal{H}$ is monotone and $1/\beta$-Lipschitzian; $0 < \gamma_n < \beta$.

– Iterate

$$\left\lfloor \begin{array}{lcll} y_n & = & x_n - \gamma B x_n & \text{(forward step)} \\ p_n & = & J_{\gamma A}\, y_n & \text{(backward step)} \\ q_n & = & p_n - \gamma B p_n & \text{(forward step)} \\ x_{n+1} & = & x_n - y_n + q_n & \end{array} \right.$$

Then $x_n \rightharpoonup z \in \operatorname{zer}(A+B)$ [9].

The open question:

- Existing splitting methods are instances of the 3 basic splitting schemes: despite their apparent complexity and disparity, can be reduced, through a variety of techniques (by duality, by defining suitable product spaces, with suitable renorming etc.) to one of them [4, 5].
- In some very special cases, it is possible to devise methods which cannot be reduced to a 2-operator scheme, for instance if $\operatorname{zer}(A) \cap \operatorname{zer}(B) \cap \operatorname{zer}(C) \neq \emptyset$, iterate

  $$x_{n+1} = (J_A \circ J_B \circ J_C) x_n \rightharpoonup z \in \operatorname{zer}(A+B+C).$$

- ***Open question:*** Can we devise a genuine (not reducible to a 2-operator scheme through some reformulation or transformation) splitting scheme for $m > 2$?
- Some reasons why the answer may be negative:
  - $\in$ is a binary relation.
  - $2 \neq 3$: it many instances (and in several open problems presented in this conference) the behavior for $m > 2$ is quite different from the behavior for $m = 2$; see, e.g., [1].

  In this case how to prove that no genuine method exists to split $m > 2$ operators?

References

[1] J.-B. Baillon, P. L. Combettes, and R. Cominetti, There is no variational characterization of the cycles in the method of periodic projections, *J. Funct. Anal.,* vol. 262, pp. 400–408, 2012.

[2] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces.* Springer, New York, 2011.

[3] H. Brézis and P.-L. Lions, Produits infinis de résolvantes, *Israel J. Math.,* vol. 29, pp. 329-345, 1978.

[4] P. L. Combettes, Systems of structured monotone inclusions: Duality, algorithms, and applications, *SIAM Journal on Optimization,* to appear.

[5] P. L. Combettes and B. C. Vũ, Variable metric forward-backward splitting with applications to monotone inclusions in duality, *Optimization,* 2013.

[6] P. L. Lions and B. Mercier, Splitting algorithms for the sum of two nonlinear operators, *SIAM J. Numer. Anal.*, vol. 16, pp. 964–979, 1979.

[7] B. Mercier, *Topics in Finite Element Solution of Elliptic Problems* (Lectures on Mathematics, no. 63). Tata Institute of Fundamental Research, Bombay, 1979.

[8] B. F. Svaiter, On weak convergence of the Douglas-Rachford method, *SIAM J. Control Optim.*, vol. 49, pp. 280–287, 2011.

[9] P. Tseng, A modified forward-backward splitting method for maximal monotone mappings, *SIAM J. Control Optim.,* vol. 38, pp. 431–446, 2000.

Entropic convexity of expected utilities

Roberto Cominetti*

Consider the map $f : [0,1]^n \to \mathbb{R}$ defined by

$$f(x) \triangleq \mathbb{E}\left[\Phi(X_1 + \cdots + X_n)\right]$$

where $\Phi : \mathbb{N} \to \mathbb{R}$ is a (discrete) convex function and the X_i's are independent Bernoulli random variables with $\mathbb{P}(X_i = 1) = x_i$. Note that $f(x)$ is a polynomial of degree n, affine with respect to each variable x_i separately. Consider the increasing sequence $c_u = \Phi(u) - \Phi(u-1)$ and let

$$\delta = \max_{u=2,\ldots,n-1} [c_{u+1} - c_u] = \max_{u=2,\ldots,n-1} [\Phi(u+1) - 2\Phi(u) + \Phi(u-1)].$$

Problem 1: Find necessary/sufficient conditions on Φ for f to be convex.

Problem 2: Prove or disprove: the entropically perturbed map

$$\tilde{f}(x) \triangleq f(x) + \textstyle\sum_{i=1}^n x_i \ln x_i + (1-x_i)\ln(1-x_i)$$

is convex as long as $\delta \leq 2$.

Problem 3: Study the bifurcation of critical points of $\tilde{f}$ as a function of δ.

Why? The critical points of $\tilde{f}$ characterize the equilibria of the repeated game dynamics studied in *"A payoff-based learning procedure and its application to traffic games"*, Games & Economic Behavior 70 (2010). The following facts are known:

a) If $\Phi(u) = a + bu$ then $f(x) = a + b\sum_{i=1}^n x_i$ is linear, hence convex.

b) If $\Phi(u) = a + bu + cu^2$ then $\tilde{f}$ is convex for $\delta \leq 2$ (*cf.* Prop. 13).

c) In general $\tilde{f}$ is known to be convex if $\delta \leq 1$ (*cf.* Prop. 12).

d) $\tilde{f}$ has a unique critical point if $\delta < 2$ (*cf.* Thm. 10).

*Departamento de Ingeniería Industrial, Universidad de Chile ⟨rccc@dii.uchile.cl⟩.

Stealing strategies

Dick Lipton asks a question about repeated games that could be interesting to discuss during the workshop. It is mentioned in the entry *Stealing Strategies* of his blog *Gödel's Lost Letter and P=NP*.

In game theory an important problem is the existence of a winning strategy, say in repeated two player games. These strategies are functions of game configurations to actions. Wether these functions are computable efficiently has not been adressed very often I think. It is however a practical requirement.

So consider as an example the game CHOMP. Two player game, who play in turns. The game starts with a chocolate bar in form of a rectangle. Later during the game the chocolate is reduced to some *Young tableau*. This is just a subset of the grid $\mathbb{N}^2$, with the convexity property, that if cell (a, b) belongs to the tableau, then any cell (a', b') with $1 \leq a' \leq a$ and $1 \leq b' \leq b$ also belongs to the tableau. Each turn one of the players selects some cell (a, b) from the tableau, and removes all cells (a', b') with $a' \geq a$ and $b' \geq b$. A player looses if it is his turn and the tableau consists of a single cell.

In this context a strategy is a function mapping a tableau to a cell from this tableau.

Say we have two players called white and black, and white starts. David Gale showed that there is a strategy for the white player such that she can win, no matter how black plays, except if white is to start on the single cell tableau of course. The argument is based on a technique called *strategy stealing*. Here is the proof. Suppose that there is a strategy for black, such that no matter what white plays, black can always win. So if the white player just removes to top right cell of the initial rectangle, then the black will have some response (a, b) to the resulting configuration. But white could play this same move (a, b) is his first move, and therefore take the role of black and win.

This shows that there is a winning strategy, without explicitly describing one. But there is a way for the white player to learn this strategy. Suppose the players play sequentially an infinite number of games. Then there is a possibility for white to win always but a finite number of times. The strategies are discrete objects which can be numbered sequentially $s_1, s_2, \ldots$. So all white has to do, is to stick to some strategy s_i starting with $i = 1$, and as soon as she looses a game she continues the next game with the strategy s_{i+1}. Then two things can happen, either she discovered the wining strategy, or she discovered a strategy that is winning on this infinite sequence of games.

Now restrict to strategies that are computable in polynomial time. This means that we consider strategies for which there is a program P and a constant c, such that on a tableau containing the top left cell $(1, a)$ and bottom right cell $(b, 1)$, the program P selects a cell from the tableau in time $O((a + b)^c)$.

Suppose that there exists a polynomial time winning strategy, and that black plays it. The stealing technique above makes sure that white can learn some polynomial strategy which permits her to win always but a finite number of times. But the running time of her implemented strategy might with a much larger exponent, than the one of black. This is not satisfactory. What can we change to make sure that she learns a winning strategy of the same complexity order than black?

C. Dürr

The length-bounded VPN design problem

Jannik Matuschke Britta Peis Felix Seibert

A *virtual private network* (VPN) is a connection of several private networks through a public network by means of point-to-point connections. The *VPN design problem* asks for a routing template within the public network such as to minimize the cost for the capacities required to guarantee sufficient connectivity for any possible pattern of communication between the private networks. Formally, we are given a graph $G = (V, E)$ with a set of *terminals* $W \subseteq V$ and edge costs $c \in \mathbb{R}^E$. A *routing template* is a function P that assigns a v-w-path P_{vw} to every unordered pair of vertices $v, w \in W$. A *demand pattern* is a matching D in the complete graph on the vertices W. Let $\mathcal{D}$ be the set of all demand patterns. The required capacity of edge e with respect to the routing template P is $u_e(P) := \max_{D \in \mathcal{D}} |\{\{v, w\} \in D : e \in P_{vw}\}|$. Accordingly, the cost of the routing template is $c(P) := \sum_{e \in E} c(e) u_e(P)$. Goyal, Olver, and Shepherd [1] showed that an optimal routing template can be found by computing a shortest path tree from any terminal and choosing the one with minimal cost, proving what before has been known as the *VPN conjecture*.

Length-bounded VPN design Virtual private networks are often used for real-time applications such as video conferencing. In these applications the delay of a connection must be below a certain threshold to ensure sufficient quality of service. Thus, a natural extension of the VPN design problem requires the paths between the terminals to fulfill a *length restriction*, i.e., we are given a length bound $L \in \mathbb{R}$ and edge lengths $\ell \in \mathbb{R}^E$ as additional input and every path P_{vw} of the routing template has to fulfill $\sum_{e \in P} \ell(e) \leq L$.

It is easy to see that in this case the set of feasible solutions might not contain a tree. In fact, the length-bounded VPN design problem can be reduced to the set cover problem and is hard to approximate by any factor better than logarithmic. We therefore consider (α, β)-approximations that relax the length bound by a factor of α while approximating the cost of the optimal solution to the non-relaxed problem by a factor of β. When relaxing the length bound by a factor of at least 2, we can achieve constant factor approximations.

Theorem (MPS 2013)**.** *There is a* $(2, 2)$*-approximation algorithm and an* $(8, 1)$*-approximation algorithm for the length-bounded VPN design problem.*

Question: Is there a $(2, 1)$-approximation algorithm for length-bounded VPN design that returns a solution that is a tree?

References

[1] Navin Goyal, Neil Olver, and F. Bruce Shepherd. The VPN conjecture is true. *Journal of the ACM*, 60(3):17, 2013.

CONVERGENCE OF INERTIAL GRADIENT FLOWS IN CONVEX OPTIMIZATION PROBLEMS

RIDA LARAKI AND PANAYOTIS MERTIKOPOULOS

Abstract. The method of gradient descent dates back to the 17-th century principle of energy minimization – the well known "heavy ball with friction" analogy. However, despite the ample physical intuition, if a convex optimization problem is treated as a physical system with the problem's objective playing the role of the system's potential energy, it is not clear whether the physical principle of energy minimization actually holds. Modulo some mild technical conditions, Alvarez (2000) showed that this is indeed the case if one applies Newton's law of motion to a smooth convex objective defined over $\mathbb{R}^n$; however, whether similar results extend to constrained convex optimization problems (and how) is a completely open question.

Statement of the Problem

Consider the "heavy ball with friction" incarnation of Newton's second law in $\mathbb{R}^n$:

$$\ddot{x} = -\operatorname{grad} V - \eta\dot{x}, \tag{HBF}$$

where the "potential energy" $V\colon \mathbb{R}^n \to \mathbb{R}$ is a smooth convex function and $\eta > 0$ is a friction coefficient which dampens the system and controls the rate of energy dissipation. Physical intuition suggests that the trajectories of (HBF) will be drawn to low-energy levels and, due to friction, will eventually converge to a minimizer of V. If $\arg\min V \neq \varnothing$, Alvarez (2000) showed that this "energy minimization" principle holds true: every solution trajectory of (HBF) converges to a minimizer of V.

Consider now the constrained convex optimization problem:

$$\begin{aligned} &\text{maximize} && V(x), \\ &\text{subject to} && x \in C, \end{aligned} \tag{P}$$

where $C \in \mathbb{R}^n$ is a compact convex set with full-dimensional interior and sufficiently nice boundary. In this case, (HBF) will hit the boundary $\operatorname{bd} C$ of C in finite time, so there is no hope of convergence. On the other hand, to counter such issues in a first order framework, Alvarez et al. (2004) introduced the Hessian Riemannian gradient system

$$\dot{x} = -\operatorname{grad}_g V \tag{HR}$$

where $(\operatorname{grad}_g V)_j = \sum_k g^{-1}_{jk} \frac{\partial V}{\partial x_k}$ denotes the *Riemannian gradient* of V w.r.t. a steep *Hessian Riemannian metric* g on C – i.e. a metric of the form $g = \operatorname{Hess}(h)$ for some strictly convex function $h \in C^\infty(\operatorname{int} C)$ with $|dh(x)| \to +\infty$ as $x \to \operatorname{bd} C$. Again, under mild technical conditions, the trajectories of (HR) converge to the minimizers of V.

The above suggests a very hopeful approach to salvage the convergence of (HBF) in constrained problems: simply take the so-called *covariant* (i.e. invariant w.r.t. parallel translations) version of Newton's law defined as:

$$\frac{D^2x}{Dt^2} = -\operatorname{grad}_g V - \eta\dot{x}, \tag{HBF$_C$}$$

where $\frac{D^2x_k}{Dt^2}$ denotes the covariant derivative operator which generalizes ordinary differentiation to a Riemannian setting – more explicitly, $\frac{D^2x_k}{Dt^2} = \ddot{x}_k + \sum_{i,j} \Gamma^k_{ij}\dot{x}_i\dot{x}_j$ where Γ^k_{ij} are the Christoffel symbols of g (Lee, 1997). We are thus led to the following open problem:

Open Problem (The Principle of Energy Minimization)**.** *Is* (HBF$_C$) *well-posed? Do the solution trajectories of* (HBF$_C$) *converge to a minimizer of* (P) *from all interior initial conditions?*

References

Alvarez, F., 2000: On the minimizing property of a second order dissipative system in Hilbert spaces. *SIAM Journal on Control and Optimization*, **38 (4)**, 1102–1119.

Alvarez, F., J. Bolte, and O. Brahic, 2004: Hessian Riemannian gradient flows in convex programming. *SIAM Journal on Control and Optimization*, **43 (2)**, 477–501.

Lee, J. M., 1997: *Riemannian Manifolds: an Introduction to Curvature*, Graduate Texts in Mathematics, Vol. 176. Springer.

Rank Aggregation with Partitioning

Alantha Newman

October 13, 2013

Given ℓ permutations $\pi_1, \pi_2, \ldots \pi_\ell$, each on the integers 1 through n, the *Rank Aggregation* problem is to find an output permutation π that minimizes the following quantity:

$$\sum_{i=1}^{\ell} d(\pi, \pi_i), \tag{1}$$

where $d(\sigma, \rho)$ is the distance between the two permutations σ and ρ. One commonly used distance measure between permutations is known as the *Kendall-Tau* or *Bubble-Sort* distance and is simply the number of inversions in one permutation when compared to the other. In other words, without loss of generality, let us assume that σ is the identity permutation. Then the distance between permutations σ and ρ is simply the number of pairs of integers in ρ that appear in the "wrong" order, i.e. a appears before b, but $a > b$. Here, we will use $d(.,.)$ to denote this distance measure.

The problem of Rank Aggregation has been well studied and there is a PTAS [KMS07]. It is known to be NP-hard when $\ell \geq 4$ [DKNS01]. There is also a simple 2-approximation: take the best of the input permutations.

Sometimes, however, it may be the case that a single permutation can not really be used to accurately describe the input permutations. For example, suppose half of the input permutations are the identity permutation and half of the input permutations are the reverse of the indentity permutation. Then it is easy to see that any permutation is a median, because for each pair of elements, it does not matter in which order it appears in the output permutation. Moreover, for any permutation π, the value of the total distance from π to all of the input permutations is high—$\binom{n}{2}(\frac{\ell}{2})$—which is intuitively why it does not give us much information about the input permutations. In this case, it would be better to divide the permutations into two sets and give two medians, in which case the value of an optimal solution would drop to zero. This leads us to the following problem:

Rank Aggregation with k Partitions: Given ℓ input permutations on $[n]$ and an integer k, find a set of k permutations, S_k, that minimize:

$$\sum_{i=1}^{\ell} \sum_{j=1}^{k} d(\pi_i, S_k), \tag{2}$$

where the distance $d(\pi_i, S_k)$ is defined as the distance between π_i and the closest permutation in S_k.

This is actually the well-known k-median problem with the restriction that the input points are permutations. It is a "continuous" version because the medians can be chosen from the entire set of permutations and not just from the input permutations, as is the case in the "discrete" version of the problem. An optimal solution to the discrete version of this problem yields a 2-approximation to the continuous version. This is because a solution to the continuous problem clusters the input permutations around the k medians. For each of these clusters, we can just pick the best permutations from that cluster, which is 2-approximation to the objective value obtained using the k optimal medians.

Thus, we can exhaustively consider the $\binom{n}{k}$ possible sets of k medians. For each set, we assign each input permutation to the closest permutation in the set. One of these sets results in a partitioning whose value is no greater than the 2-approximation to the optimal solution. This algorithm is not efficient for large values of k since it runs in time $O(n^k)$. (Note that we could also use this approach to find the best partitioning with at most k medians in the same asymptotic running time.) Thus, we have the following open problems:

Problem 1 *Give a 2-approximation for the Rank Aggregation with k Partitions problem that has a polynomial running time (independent of k).*

Problem 2 *Give an optimal algorithm for the discrete Rank Aggregation with k Partitions problem that has a polynomial running time (independent of k).*

Problem 3 *Give any approximation for the Rank Aggregation with k Partitions problem that has polynomial running time and has approximation guarantee better than $1+\sqrt{3}$, which is the current best for general k-median.*

Note that a solution for Problem 2 yields a solution for Problem 1. However, it not clear how the problems are related in the other direction.

References

[DKNS01] Cynthia Dwork, Ravi Kumar, Moni Naor, and Dandapani Sivakumar. Rank aggregation methods for the web. In *Proceedings of the 10th international conference on World Wide Web*, pages 613–622. ACM, 2001.

[KMS07] Claire Kenyon-Mathieu and Warren Schudy. How to rank with few errors. In David S. Johnson and Uriel Feige, editors, *STOC*, pages 95–103. ACM, 2007.

The connectivity problem in the *number-in-hand* computation model

Ivan Rapaport

Departamento de Ingeniería Matemática, Universidad de Chile

Suppose that we have a network of n processors where each processor knows its own ID and the IDs of its neighbors. Processors communicate in synchronous rounds by writing messages on a whiteboard, which is visible to all of them. The goal is to design a protocol at the end of which every processor knows whether the network is connected. We assume the following:

- Each processor in the network has a different ID (a number between 1 and n, where n is the number of processors in the network).
- The only information each processor has, besides n and its own ID, is the list of IDs of its neighboring processors.
- Processors have unlimited computational power.
- When the protocol ends every processor should know whether the nework is connected.

We are interested in two complexity measures:

1. Number of rounds (where in each round all processors write simultaneously one message on the whiteboard).

2. Message size (number of bits of the longest message written during the process on the whiteboard).

If there is no restriction in the message size then there is a trivial one-round protocol that reconstructs any network. In fact, given an arbitrary network G (simple undirected graph) and given an arbitrary assignment of IDs to the n processors (nodes of G) the protocol is the following: every processor writes on the whiteboard the $0-1$ vector $x \in \{0,1\}^n$ corresponding to the indicator function of its neighborhood. With this information written on the whiteboard every processor can reconstruct G.

We conjecture the following: there is no $O(\log n)$ message size protocol that solves the connectivity problem in one round.

Lagrangian duality in Online Scheduling

Nguyen Kim Thang

The most powerful tool until now to design and analyze online algorithms is the potential function methods. However, the construction of the algorithms is far from trivial and that gives little insight about the nature of the problems and the algorithms. Recently, principled methods based on non-linear mathematical programming have been introduced to study online algorithms [1, 2, 3]. The idea of the approaches is the following.

Consider a mathematical programming relaxation (associated with a given problem) which is not necessarily convex and its Lagrangian dual. Then construct dual variables such that the Lagrangian dual has objective value within a desired factor of the primal one (due to some algorithm). Then by the standard Lagrangian weak duality for mathematical programming, the competitive ratio follows. The main step is the construction of dual variables (dual fitting or primal dual); and subsequently a competitive algorithm could be derived from such dual variables.

With the unified approaches, algorithms have been designed simpler and have better performance guarantee [3]. We are interested in studying the following problem.

Problem We have one machine and a energy budget B. The consumed energy is $\int_0^\infty P(s(t))$ given $s(t)$ the speed of machine at time t and P is a convex function; a standard energy power function $P(s) = s^\alpha$ where α is a constant. Jobs arrive over time, job j have value v_j, processing time p_j, release date and deadline are r_j and d_j, respectively. The objective is to maximize the total value of jobs completed on time without violating the energy budget.

Question Does there exist a $(1+\epsilon)$-resource augmentation, $O(1/\epsilon^k)$-competitive algorithm for the problem where k is some constant?

We say that an algorithm is α-resource augmentation, r-competitive if the total value of the algorithm (with the energy power function $P(s)/\alpha$) is at least factor r from the total value of the optimal algorithm (with the energy power function $P(s)$).

References

[1] S. Anand, Naveen Garg, and Amit Kumar. Resource augmentation for weighted flow-time explained by dual fitting. In *Proc. 23rd ACM-SIAM Symposium on Discrete Algorithms*, pages 1228–1241, 2012.

[2] Anupam Gupta, Ravishankar Krishnaswamy, and Kirk Pruhs. Online primal-dual for non-linear optimization with applications to speed scaling. In *Proc. 10th Workshop on Approximation and Online Algorithms*, pages 173–186, 2012.

[3] Nguyen Kim Thang. Lagrangian duality in online scheduling with resource augmentation and speed scaling. In *Proc. 21st European Symposium on Algorithms*, 2013.

Consistency of decisions at different time horizons

Jorge R. Vera

October, 2013

Consider the following situation: We have to make decisions (for instance, production decisions in a company) which cover, say, a year divided in months. There are costs associated to those decisions as well as various kind of constraints. This will be the typical case of a production planning problem where we have to decide on monthly production quantities as well as on production resources. Now, consider the analogous short-term operational planning problem: suppose we are at the beginning of the first month and we have to decide on weekly (or even daily) production quantities, using available resources, some of which have been already "fixed" in the planning step. This situation is common in practice and, due to uncertainties and different degrees of model aggregation, most of the time inconsistencies appear which could lead to excessive cost or even indefeasibilities.

The problem is how to make good tactical decisions, taking into account the effect in the short term problem.

This can be represented as one problem at the tactical level, of the form TP) : $\min\{c^T x : Ax = b, x \geq 0\}$, where x is a vector of monthly variables, and a problem for, say, the first period: OP) : $q(x, \xi) = \min\{d^T y : Tx + Wy = h, y \geq 0\}$ where $\xi = (d, T, W, h)$ is the data vector which is random in some of its components and can be used to represent the fact that the actual situation in the short term will be different from the one assumed in the tactical horizon.

Questions:

1. How can we make decisions in TP) in such a way that there is a high probability that OP) is feasible?
2. How can we make decisions in TP) in such a way that OP) does not deviate excessively (to be defined) from the optimal consistent situation (say, when there is no uncertainty).
3. One way to model the interaction is as a Two Stage Stochastic problem of the form

$$\begin{array}{ll} \min & c^T x + Q(x) \\ s.a. & Ax = b, x \geq 0 \end{array}$$

   where $Q(x) = E(q(x, \xi))$. In this case, we want to consider some recent algorithmic alternatives which have emerged from recent work in First Order Methods. For instance, the Stochastic Subgradient Method can be applied in an accelerated way to this problem, and recent developments by G. Lan suggest that a randomized version of Conditional Gradient could be competitive for solving this problem.
4. In any of the above cases, we want to have estimates of the probability of consistency.