Convex optimization: applications, formulations, relaxations

Jérôme MALICK

CNRS, LJK/INRIA, BiPoP team

NECS team seminar (INRIA/GIPSA) – September 28, 2011

Convex optimization, useful applied maths

Optimization in two words:

- "the maths of doing-better" or "the maths of decision-making"
- Mature disciplin of applied maths (theory, algorithms, software)
- Recent explosion of applications in engineering sciences

Convex optimization, useful applied maths

Optimization in two words:

- “the maths of doing-better” or “the maths of decision-making”
- Mature disciplin of applied maths (theory, algorithms, software)
- Recent explosion of applications in engineering sciences

And why convex ?... because it's the favorable case !

"the great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity" T. Rockafellar

- Geometrical properties : globality, guarantees,...
- Useful tools: duality, sensitivity analysis... and algorithms !

Convex optimization, useful applied maths

Optimization in two words:

- "the maths of doing-better" or "the maths of decision-making"
- Mature disciplin of applied maths (theory, algorithms, software)
- Recent explosion of applications in engineering sciences

And why convex ?... because it's the favorable case !

"the great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity" T. Rockafellar

- Geometrical properties : globality, guarantees,...
- Useful tools: duality, sensitivity analysis... and algorithms !

Some current trends and challenges in (convex) optimization

1. more complex models ("nonlinear mixed-integer programming")
2. very large-scale data ("huge-scale optimization")
3. noisy or unknown data ("robust optimization")
4. "applied optimization" ("incremental optimization")

Optimization: BiPoP... in fact, me :-)

Projects that I have been leading (since I arrived in nov. 2006)

- Algorithms for optimisation convex and...
  - ... differentiable (with M. Fuentes, ancien post-doc)
  - ... differentiable on submanifold (with P.-A. Absil, Belgium)
  - ... nonsmooth (with A. Daniilidis, Barcelona)
  - ... semidefinite (with F. Rendl, Austria)
- Industrial applications:
  - electrical production (with C. Lemaréchal and EDF)
  - finance (with RaisePartner)
- Applications in other disciplins :
  - combinatorial optimisation (with F. Roupin, Paris)
  - polynomial optimisation (with D. Henrion, LAAS)
  - computational mechanics (with V. Acary and Schneider)
  - statistical learning (with Lear and Yahoo!)
- and some related theoretical questions... in particular:
  - properties of spectral functions (with H. Sendov, Canada)
  - convergence of projection algorithms (with A. Lewis, US)

Optimization: BiPoP... in fact, me :-)

Projects that I have been leading (since I arrived in nov. 2006)

- Algorithms for optimisation convex and...
  - ... differentiable (with M. Fuentes, ancien post-doc)
  - ... differentiable on submanifold (with P.-A. Absil, Belgium)
  - ... nonsmooth (with A. Daniilidis, Barcelona)
  - ... semidefinite (with F. Rendl, Austria)
- Industrial applications:
  - electrical production (with C. Lemaréchal and EDF)
  - finance (with RaisePartner)
- Applications in other disciplins :
  - combinatorial optimisation (with F. Roupin, Paris)
  - polynomial optimisation (with D. Henrion, LAAS)
  - computational mechanics (with V. Acary and Schneider)
  - statistical learning (with Lear and Yahoo!)
- and some related theoretical questions... in particular:
  - properties of spectral functions (with H. Sendov, Canada)
  - convergence of projection algorithms (with A. Lewis, US)

Today's presentation

Roadmap

1. Optimization of the French electricity production (20 years)
   Bipop (C. Lemaréchal, J. Malick, S. Zaourar) and EDF
2. Low-rank algorithm for multiclass classification (1 year)
   Lear (Z. Harchaoui), Bipop (J. Malick) and Yahoo! (M. Dudik)
3. New SDP/LMI relaxations in combinatorial optimization (3 years)
   Bipop (N. Krislock, J. Malick) and Paris 13 (F. Roupin)

"Goals" (modest)

- show problems, issues, ideas - with no details
- for you : give a (partial) overview (of what happens upstairs)
- for me : insisting on my hot topics... and do local advertising

Examples of applications of convex optimization

1. **Optimization of electricity production**
2. Low-rank penalization for multiclass classification
3. New SDP relaxations in combinatorial optimization

Short-term electricity production managment

In France: electricity produced by $n \simeq 200$ production units

nuclear 80%

oil + coal 3%

water 17%

Short-term electricity production managment

In France: electricity produced by $n \simeq 200$ production units

nuclear 80% | oil + coal 3% | water 17%

Hard optimization problem: large-scale, heterogeneous, short time

$$\begin{cases} \min \quad \sum_i c_i(p_i) & \text{(sum of costs)} \\ \quad p_i \in P_i \quad i = 1, \ldots, n & \text{(technical constraints)} \\ \quad \sum_i p_i^t = d^t \quad t = 1, \ldots, T & \text{(answer to demand)} \end{cases}$$

Short-term electricity production managment

In France: electricity produced by $n \simeq 200$ production units

nuclear 80%

oil + coal 3%

water 17%

Hard optimization problem: large-scale, heterogeneous, short time

$$\begin{cases} \min \quad \sum_i c_i(p_i) & \text{(sum of costs)} \\ \quad p_i \in P_i \quad i = 1, \dots, n & \text{(technical constraints)} \\ \quad \sum_i p_i^t = d^t \quad t = 1, \dots, T & \text{(answer to demand)} \quad \leftarrow \lambda^t \end{cases}$$

Decomposable by duality !

$$L(p, \lambda) \;:=\; \sum_i c_i(p_i) + \sum_t \lambda^t \Big(\sum_i p_i^t - d^t\Big) = \sum_i \Big(c_i(p_i) + \sum_t \lambda^t (p_i^t - d^t)\Big)$$

Short-term electricity production managment

In France: electricity produced by $n \simeq 200$ production units

nuclear 80%

oil + coal 3%

water 17%

Hard optimization problem: large-scale, heterogeneous, short time

$$\begin{cases} \min \quad \sum_i c_i(p_i) & \text{(sum of costs)} \\ \quad p_i \in P_i \quad i = 1, \dots, n & \text{(technical constraints)} \\ \quad \sum_i p_i^t = d^t \quad t = 1, \dots, T & \text{(answer to demand)} \quad \leftarrow \lambda^t \end{cases}$$

Decomposable by duality !

$$L(p, \lambda) := \sum_i c_i(p_i) + \sum_t \lambda^t \Big(\sum_i p_i^t - d^t \Big) = \sum_i \Big(c_i(p_i) + \sum_t \lambda^t (p_i^t - d^t) \Big)$$

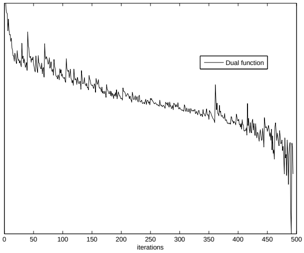
$$\theta(\lambda) := \min_{p \in P} L(p, \lambda) = \sum_i \min_{p_i \in P_i} \Big(c_i(p_i) + \sum_t \lambda^t (p_i^t - d^t) \Big)$$

Centralized resolution by convex optimization

- Computing $\theta(\lambda) =$ solving the n independent sub-problems
  - thermic units: dynamic optimization
  - hydrolic valleys : mixed-integer linear programming (CPLEX)

Centralized resolution by convex optimization

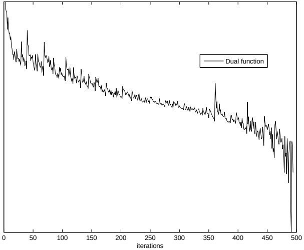
- Computing $\theta(\lambda)$ = solving the n independent sub-problems
  - thermic units: dynamic optimization
  - hydrolic valleys : mixed-integer linear programming (CPLEX)
- Maximizing concave nonsmooth function θ
  - gives optimal prices $\lambda^\star \in \mathbb{R}^T$
  - initializes a 2nd phase heuristic that computes feasible plannings p_i
- E.g.: production 35000 - 70000 MW, mismatch at most 30MW

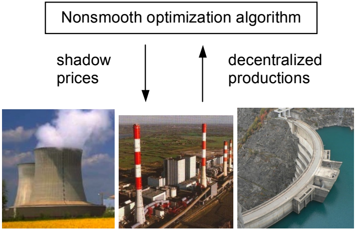


Centralized resolution by convex optimization

- Computing $\theta(\lambda)$ = solving the n independent sub-problems
  - thermic units: dynamic optimization
  - hydrolic valleys : mixed-integer linear programming (CPLEX)

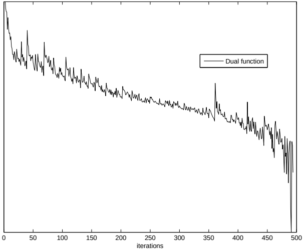
- Maximizing concave nonsmooth function θ
  - gives optimal prices $\lambda^\star \in \mathbb{R}^T$
  - initializes a 2nd phase heuristic that computes feasible plannings p_i
- E.g.: production 35000 - 70000 MW, mismatch at most 30MW

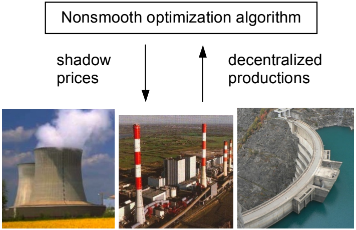




Centralized resolution by convex optimization

- Computing $\theta(\lambda)$ = solving the n independent sub-problems
  - thermic units: dynamic optimization
  - hydrolic valleys : mixed-integer linear programming (CPLEX)
- Maximizing concave nonsmooth function θ
  - gives optimal prices $\lambda^\star \in \mathbb{R}^T$
  - initializes a 2nd phase heuristic that computes feasible plannings p_i
- E.g.: production 35000 - 70000 MW, mismatch at most 30MW





- On-going research: **1)** handle noise, stabilisation of prices
  2) interpretation of prices, new modeling

Examples of applications of convex optimization

1. Optimization of electricity production
2. **Low-rank penalization for multiclass classification**
3. New SDP relaxations in combinatorial optimization

Supervized multiclass classification

- Goal : classify objets (associated features)
  - data : couples feature/class $(x_i, y_i) \in \mathbb{R}^p \times \{0,1\}^K$
  - assign a class to a new incoming object described by x ?
- E.g. : biostats - computer vision

Supervized multiclass classification

- Goal : classify objets (associated features)
  - data : couples feature/class $(x_i, y_i) \in \mathbb{R}^p \times \{0,1\}^K$
  - assign a class to a new incoming object described by x ?
- E.g. : biostats - computer vision
- Optimization: learning a classifier from data
  - compute a "weight" matrix $W \in \mathbb{R}^{p \times K}$
  - minimize an error function (+ regularization)

$$\min_{W \in \mathbb{R}^{p \times K}} \frac{1}{n} \sum_{i=1}^{n} L(y_i, W^\top x_i) + \alpha \operatorname{Reg}(W)$$

  - classify as $\max_{k=1,\ldots,K} {w_k}^\top x$ (with w_k columns of W)

Supervized multiclass classification

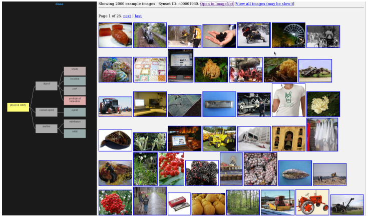
- Goal : classify objets (associated features)
  - data : couples feature/class $(x_i, y_i) \in \mathbb{R}^p \times \{0,1\}^K$
  - assign a class to a new incoming object described by x ?
- E.g. : biostats - computer vision
- Optimization: learning a classifier from data
  - compute a "weight" matrix $W \in \mathbb{R}^{p \times K}$
  - minimize an error function (+ regularization)

$$\min_{W \in \mathbb{R}^{p \times K}} \frac{1}{n} \sum_{i=1}^{n} L(y_i, W^\top x_i) + \alpha \operatorname{Reg}(W)$$

  - classify as $\max_{k=1,\dots,K} {w_k}^\top x$ (with w_k columns of W)
- State of the art: OK for $K = 2$ (and $K \approx 10$)
- All this is well-known, but the challenge is : K large
- Eg: Pascal Challenge '10 "imagenet" : $K \approx 1000$ (but not reliable)

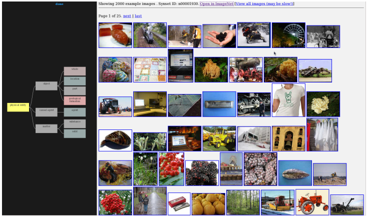
Multiclass classification with sub-structure

- If unknown underlying structure...
- $\rightarrow$ Low rank W
- $W = UV^\top,\ U \in \mathbb{R}^{p \times r}, V \in \mathbb{R}^{K \times r}$
- Factorization of information
  (and cheaper computations)



Multiclass classification with sub-structure

- If unknown underlying structure...
- $\rightarrow$ Low rank W
- $W = UV^{\top}, \; U \in \mathbb{R}^{p \times r}, V \in \mathbb{R}^{K \times r}$
- Factorization of information
  (and cheaper computations)

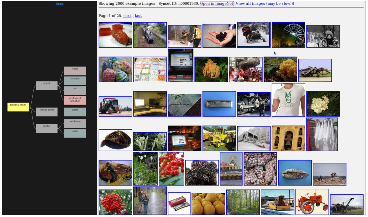


- Optimization: large-scale non-convex problem

$$\min_{W \in \mathbb{R}^{p \times K}} \; \alpha \, \mathrm{rank}(W) \; + \frac{1}{n} \sum_{i=1}^{n} L(y_i, W^{\top} x_i)$$

Multiclass classification with sub-structure

- If unknown underlying structure...
- $\rightarrow$ Low rank W
- $W = UV^\top,\ U \in \mathbb{R}^{p \times r}, V \in \mathbb{R}^{K \times r}$
- Factorization of information
  (and cheaper computations)



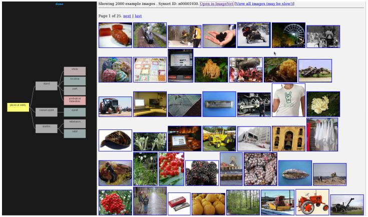
- Optimization: nonsmooth convex relaxation

$$\min_{W \in \mathbb{R}^{p \times K}} \ \alpha \, \|\sigma(W)\|_1 \ + \frac{1}{n} \sum_{i=1}^{n} L(y_i, W^\top x_i)$$

where $\sigma(W)$ is the vector of singular values

Multiclass classification with sub-structure

- If unknown underlying structure...
- $\rightarrow$ Low rank W
- $W = UV^\top,\ U \in \mathbb{R}^{p \times r}, V \in \mathbb{R}^{K \times r}$
- Factorization of information
  (and cheaper computations)



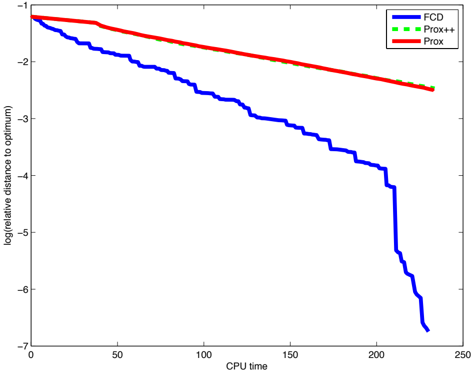
- Optimization: nonsmooth convex relaxation

$$\min_{W \in \mathbb{R}^{p \times K}} \ \alpha \, \|\sigma(W)\|_1 \ + \frac{1}{n} \sum_{i=1}^{n} L(y_i, W^\top x_i)$$

  where $\sigma(W)$ is the vector of singular values
- Our approach: constructive, rang control, no SVD
  (handle only rank-one matrices)
- Cheap algorithm of type "greedy coordinate descent"

Numerical illustration

- 2010 Pascal Visual Object Classes Challenge ("imagenet")
- $K = 1000$ $(p = 600)$
- Plot: relative error $\log\left((f(x_k) - f^*)/f^*\right)$ vs time
- Our algorithm outperforms the state-of-the-art



Examples of applications of convex optimization

1. Optimization of electricity production
2. Low-rank penalization for multiclass classification
3. **New SDP relaxations in combinatorial optimization**

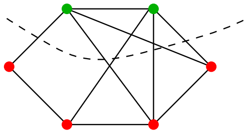
Combinatorial optimization

- Many combinatorial optimization problems can be written as quadratic problems under quadratic constraints

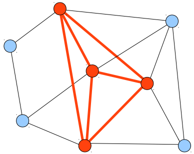
$$(\mathrm{P}) \quad \begin{cases} \min \quad x^\top Q_0\, x + {b_0}^\top x \\ \quad x^\top Q_i\, x + {b_i}^\top x + c_i = (\text{or} \leqslant)\, 0 \\ \quad x \in \{0,1\}^n \end{cases}$$

- Examples: (in graphs, models for real-life applications)

find a max cut



find a densest k-subgraph



- Problems: NP-hard... and, in fact, hard to solve (eg: $n = 100$)
- Tool: smart enumeration (branch-and-bound) using lower bounds of the optimal value of (P) (and sub-problems of (P))

Convex relaxations

- "Relaxation" : relax some constraints of (P) to get a convex problem (R) that we solve "easily"

$$\mathrm{val(P)} \geqslant \mathrm{val(R)}$$

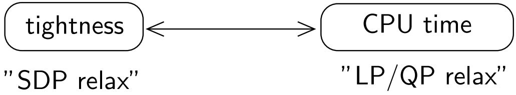
- Eg: replace $\{0, 1\}$ by $[0, 1]$ $\rightarrow$ linear problems (CPLEX)

Convex relaxations

- “Relaxation” : relax some constraints of (P) to get a convex problem (R) that we solve “easily”

$$\text{val(P)} \geqslant \text{val(R)}$$

- Eg: replace $\{0, 1\}$ by $[0, 1]$ $\rightarrow$ linear problems (CPLEX)
- Quality of the relaxation: balance between tightness and computing time

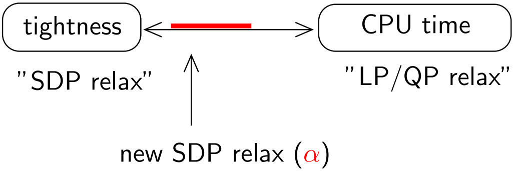


Convex relaxations

- “Relaxation” : relax some constraints of (P) to get a convex problem (R) that we solve “easily”

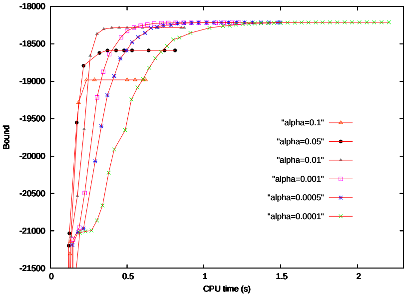
$$\text{val(P)} \geqslant \text{val(R)}$$

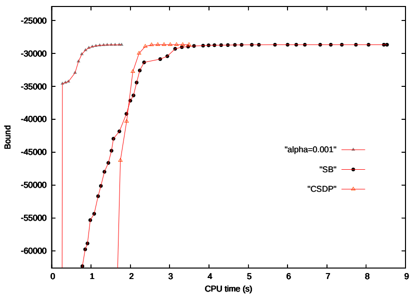
- Eg: replace $\{0,1\}$ by $[0,1]$ $\rightarrow$ linear problems (CPLEX)
- Quality of the relaxation: balance between tightness and computing time



- New relaxations of type SDP (ou LMI)
  - goal: have (R) solved quickly
  - strategy: have SDP quality without its price (good pruning in B&B)
  - technique: original reformulation + convex duality

Numerical illustrations





- New bounds and solver allow us
  - to attain LP/QP bounds quicker than with dedicated solvers
  - to (almost) attain SDP bounds quicker than with dedicated solvers

- Densest k-subgraph: a simple exact resolution scheme using these bounds outperforms the best existing approach (using CPLEX) (due to a better prunning in the B&B search tree...)