Towards Stronger Relaxations for ℓ_0-Regularized Problems

Théo Guyard
Inria and Insa Rennes

Workshop on ℓ_0-based minimization
Institut Henri Poincaré

Joint work with Cédric Herzet, Clément Elvira, Ayşe-Nur Arslan
and discussions with Emmanuel Soubies

Problem of Interest

ℓ_0-Regularized Problems

$$p^\star = \min_{\mathbf{x} \in \mathbf{R}^n} \; f(\mathbf{x}) + \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$$

Problem of Interest

ℓ_0-Regularized Problems

$$p^\star = \min_{\mathbf{x} \in \mathbf{R}^n} \; f(\mathbf{x}) + \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$$

Main Question

$$\boxed{\text{How to construct a generic Branch-and-Bound algorithm?}}$$

Problem of Interest

ℓ_0-Regularized Problems

$$p^\star = \min_{\mathbf{x} \in \mathbf{R}^n} \; f(\mathbf{x}) + \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$$

Main Question

How to construct a generic Branch-and-Bound algorithm?

$\downarrow$

How to construct a generic relaxation method?

Problem of Interest

ℓ_0-Regularized Problems

$$p^\star = \min_{\mathbf{x} \in \mathbf{R}^n} \; f(\mathbf{x}) + \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$$

Main Question

How to construct a generic Branch-and-Bound algorithm?

$\downarrow$

How to construct a generic relaxation method?

Framework:

- f is proper, closed, convex
- h is proper, closed, convex and separable as $h = \sum_i h_i$
- h_i are supercoercive, even, and verify $h_i \geq h_i(0) = 0$
- $\mathbf{0} \in \operatorname{int}\operatorname{dom}(f) \cap \operatorname{int}\operatorname{dom}(h)$

Problem of Interest

ℓ_0-Regularized Problems

$$p^\star = \min_{\mathbf{x} \in \mathbf{R}^n} \; f(\mathbf{x}) + \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$$

Main Question

How to construct a generic Branch-and-Bound algorithm?

$\downarrow$

How to construct a generic relaxation method?

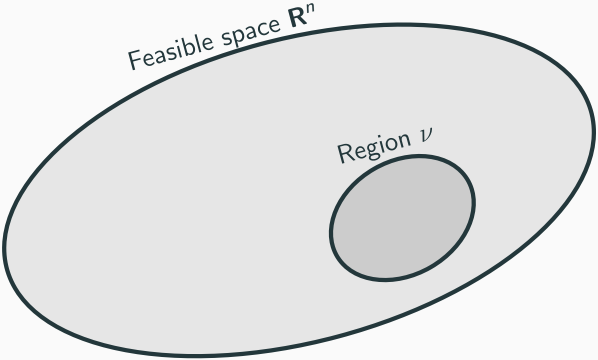
Framework:

- f is proper, closed, convex
- h is proper, closed, convex and separable as $h = \sum_i h_i$
- h_i are supercoercive, even, and verify $h_i \geq h_i(0) = 0$
- $\mathbf{0} \in \operatorname{int} \operatorname{dom}(f) \cap \operatorname{int} \operatorname{dom}(h)$

Branch-and-Bound Ingredients

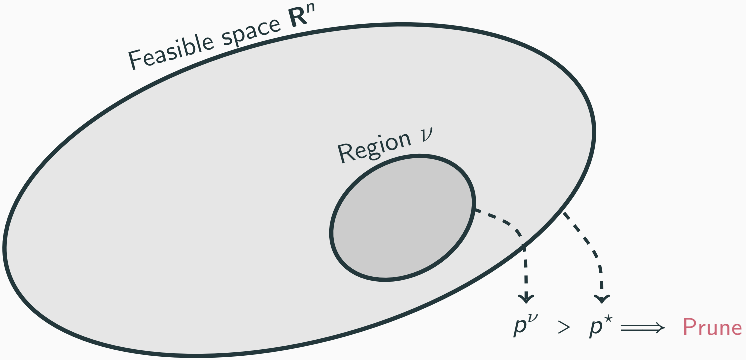
Branch-and-Bound Ingredients

Principle: Explore regions in the feasible space and prune those that cannot contain any optimal solution.



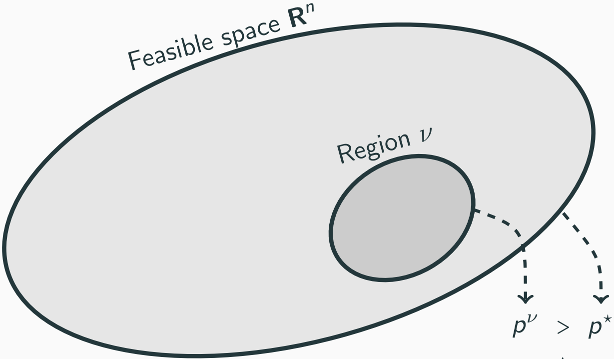
Branch-and-Bound Ingredients

Principle: Explore regions in the feasible space and prune those that cannot contain any optimal solution.



Branch-and-Bound Ingredients

Principle: Explore regions in the feasible space and prune those that cannot contain any optimal solution.



$$p^\nu \quad > \quad p^\star \implies \text{Prune}$$

$$\Uparrow$$

$$p^\nu \geq p^\nu_{\mathrm{lb}} > p^\star_{\mathrm{ub}} \geq p^\star$$

Pruning test

Branch-and-Bound Ingredients

Region: $\nu = (\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_\bullet) \equiv$ partition of $\{1, \ldots, n\}$ driving sparsity:

- $x_i = 0$ for all $i \in \mathcal{S}_0$
- $x_i \neq 0$ for all $i \in \mathcal{S}_1$
- $x_i \in \mathbf{R}$ for all $i \in \mathcal{S}_\bullet$

Branch-and-Bound Ingredients

Region: $\nu = (\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_\bullet) \equiv$ partition of $\{1, \dots, n\}$ driving sparsity:

- $x_i = 0$ for all $i \in \mathcal{S}_0$
- $x_i \neq 0$ for all $i \in \mathcal{S}_1$
- $x_i \in \mathbf{R}$ for all $i \in \mathcal{S}_\bullet$

$$p^\star = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$$

$\downarrow$ constrainted to region ν

$$p^\nu = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + g^\nu(\mathbf{x})$$

Branch-and-Bound Ingredients

Region: $\nu = (\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_\bullet) \equiv$ partition of $\{1, \ldots, n\}$ driving sparsity:

- $x_i = 0$ for all $i \in \mathcal{S}_0$
- $x_i \neq 0$ for all $i \in \mathcal{S}_1$
- $x_i \in \mathbf{R}$ for all $i \in \mathcal{S}_\bullet$

$$p^\star \quad = \min_{\mathbf{x} \in \mathbf{R}^n} \; f(\mathbf{x}) + \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$$

$\downarrow$ constrainted to region ν

$$p^\nu \quad = \min_{\mathbf{x} \in \mathbf{R}^n} \; f(\mathbf{x}) + g^\nu(\mathbf{x}) \quad \text{with} \quad g_i^\nu(x_i) = \begin{cases} \mathbf{I}(x_i = 0) & \text{if } i \in \mathcal{S}_0 \\ h_i(x_i) + \lambda & \text{if } i \in \mathcal{S}_1 \\ h_i(x_i) + \lambda \|x_i\|_0 & \text{if } i \in \mathcal{S}_\bullet \end{cases}$$

Branch-and-Bound Ingredients

Region: $\nu = (\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_\bullet) \equiv$ partition of $\{1, \dots, n\}$ driving sparsity:

- $x_i = 0$ for all $i \in \mathcal{S}_0$
- $x_i \neq 0$ for all $i \in \mathcal{S}_1$
- $x_i \in \mathbf{R}$ for all $i \in \mathcal{S}_\bullet$

$$p^\star \quad = \min_{\mathbf{x} \in \mathbf{R}^n} \; f(\mathbf{x}) + \lambda \|\mathbf{x}\|_0 + h(\mathbf{x})$$

$\downarrow$ constrainted to region ν

$$p^\nu \quad = \min_{\mathbf{x} \in \mathbf{R}^n} \; f(\mathbf{x}) + g^\nu(\mathbf{x}) \quad \text{with} \quad g_i^\nu(x_i) = \begin{cases} \mathbf{I}(x_i = 0) & \text{if } i \in \mathcal{S}_0 \\ h_i(x_i) + \lambda & \text{if } i \in \mathcal{S}_1 \\ h_i(x_i) + \lambda \|x_i\|_0 & \text{if } i \in \mathcal{S}_\bullet \end{cases}$$

$\geq$

$p^\nu_{\mathrm{lb}} \quad =$ optimal value of a relaxation

Branch-and-Bound Ingredients

Region: $\nu = (\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_\bullet) \equiv$ partition of $\{1, \dots, n\}$ driving sparsity:

- $x_i = 0$ for all $i \in \mathcal{S}_0$
- $x_i \neq 0$ for all $i \in \mathcal{S}_1$
- $x_i \in \mathbf{R}$ for all $i \in \mathcal{S}_\bullet$

$$p^\star = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + \lambda\|\mathbf{x}\|_0 + h(\mathbf{x})$$

$\downarrow$ constrainted to region ν

$$p^\nu = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + g^\nu(\mathbf{x}) \quad \text{with} \quad g_i^\nu(x_i) = \begin{cases} \mathbf{I}(x_i = 0) & \text{if } i \in \mathcal{S}_0 \\ h_i(x_i) + \lambda & \text{if } i \in \mathcal{S}_1 \\ h_i(x_i) + \lambda\|x_i\|_0 & \text{if } i \in \mathcal{S}_\bullet \end{cases}$$

$\vee\!|$

p_{lb}^ν = optimal value of a relaxation

$\boxed{\text{The relaxation needs to be valid, tight and tractable}}$

Convex Relaxations

Convex Relaxations

Idea: Convexify the objective function

$$p^{\nu} \quad = \min_{\mathbf{x}\in\mathbf{R}^n} \; f(\mathbf{x}) + g^{\nu}(\mathbf{x})$$

$$\vee|$$

$$p^{\nu}_{\mathrm{cvx}}$$

Idea: Convexify the objective function

$$p^{\nu} = \min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + g^{\nu}(\mathbf{x}) \quad \text{with} \quad g_i^{\nu}(x_i) = \begin{cases} \mathbf{I}(x_i = 0) & \text{if } i \in \mathcal{S}_0 \\ h_i(x_i) + \lambda & \text{if } i \in \mathcal{S}_1 \\ h_i(x_i) + \lambda \|x_i\|_0 & \text{if } i \in \mathcal{S}_\bullet \end{cases}$$

$$\vee\!|$$

$$p^{\nu}_{\mathrm{cvx}}$$

Convex Relaxations

Idea: Convexify the objective function

$$p^{\nu} = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + g^{\nu}(\mathbf{x}) \quad \text{with} \quad g_i^{\nu}(x_i) = \begin{cases} \mathbf{I}(x_i = 0) & \text{if } i \in \mathcal{S}_0 \\ h_i(x_i) + \lambda & \text{if } i \in \mathcal{S}_1 \\ h_i(x_i) + \lambda \|x_i\|_0 & \text{if } i \in \mathcal{S}_\bullet \end{cases}$$

$$\vee\!|$$

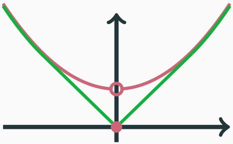
$$p^{\nu}_{\mathrm{cvx}} = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + g^{\nu}_{\mathsf{cvx}}(\mathbf{x}) \quad \text{with} \quad g^{\nu}_{\mathsf{cvx},i}(x_i) = \begin{cases} \mathbf{I}(x_i = 0) & \text{if } i \in \mathcal{S}_0 \\ h_i(x_i) + \lambda & \text{if } i \in \mathcal{S}_1 \\ \text{bi-conjugate} & \text{if } i \in \mathcal{S}_\bullet \end{cases}$$

Idea: Convexify the objective function

$$p^{\nu} = \min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + g^{\nu}(\mathbf{x}) \quad \text{with} \quad g_i^{\nu}(x_i) = \begin{cases} \mathbf{I}(x_i = 0) & \text{if } i \in \mathcal{S}_0 \\ h_i(x_i) + \lambda & \text{if } i \in \mathcal{S}_1 \\ h_i(x_i) + \lambda\|x_i\|_0 & \text{if } i \in \mathcal{S}_\bullet \end{cases}$$

$$\vee\!\mid$$

$$p^{\nu}_{\mathrm{cvx}} = \min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + g^{\nu}_{\mathrm{cvx}}(\mathbf{x}) \quad \text{with} \quad g^{\nu}_{\mathrm{cvx},i}(x_i) = \begin{cases} \mathbf{I}(x_i = 0) & \text{if } i \in \mathcal{S}_0 \\ h_i(x_i) + \lambda & \text{if } i \in \mathcal{S}_1 \\ \text{bi-conjugate} & \text{if } i \in \mathcal{S}_\bullet \end{cases}$$



Plot: Functions g_i^{ν} and $g^{\nu}_{\mathrm{cvx},i}$ when $i \in \mathcal{S}_\bullet$ for $h_i(x_i) = x_i^2$.

Spotlight Result 1

The bi-conjugate admits a generic closed-form expression.

Let $g_i(x) = h_i(x) + \lambda\|x\|_0$, we have

$$g_i^{**}(x) = \begin{cases} \tau_i|x| & \text{if } |x| \leq \mu_i \\ h_i(x) + \lambda & \text{otherwise} \end{cases}$$

where (τ_i, μ_i) are some "easy-to-compute" quantities.

Convex Relaxations

Spotlight Result 1

The bi-conjugate admits a generic closed-form expression.

Let $g_i(x) = h_i(x) + \lambda\|x\|_0$, we have

$$g_i^{**}(x) = \begin{cases} \tau_i|x| & \text{if } |x| \leq \mu_i \\ h_i(x) + \lambda & \text{otherwise} \end{cases}$$

where (τ_i, μ_i) are some "easy-to-compute" quantities.

$\downarrow$

Efficient solution methods for the convex relaxation

Proximal gradient

Coordinate descent

Splitting methods

...

Dual Relaxations

Dual Relaxations

Idea: Dualize the problem

$$p^\nu = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + g^\nu(\mathbf{x})$$

$$\vee\!\mid$$

$$p^\nu_{\text{dual}}$$

Dual Relaxations

Idea: Dualize the problem

$$p^{\nu} = \min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + g^{\nu}(\mathbf{x})$$

$$\vee|$$

$$p^{\nu}_{\mathrm{dual}} = \max_{\mathbf{x}\in\mathbf{R}^n} -f^*(-\mathbf{x}) - (g^{\nu})^*(\mathbf{x})$$

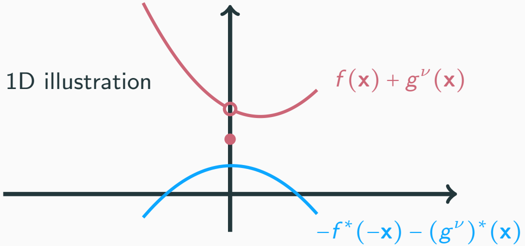
Dual Relaxations

Idea: Dualize the problem

$$p^\nu = \min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + g^\nu(\mathbf{x})$$

$$\geq$$

$$p^\nu_{\mathrm{dual}} = \max_{\mathbf{x}\in\mathbf{R}^n} -f^*(-\mathbf{x}) - (g^\nu)^*(\mathbf{x})$$



Dual Relaxations

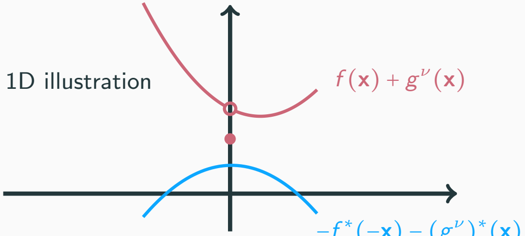
Idea: Dualize the problem

$$p^{\nu} = \min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + g^{\nu}(\mathbf{x}) \quad \text{with} \quad g_i^{\nu}(x_i) = \begin{cases} \mathbf{I}(x_i = 0) & \text{if } i \in \mathcal{S}_0 \\ h_i(x_i) + \lambda & \text{if } i \in \mathcal{S}_1 \\ h_i(x_i) + \lambda\|x_i\|_0 & \text{if } i \in \mathcal{S}_\bullet \end{cases}$$

$\vee|$

$$p^{\nu}_{\mathrm{dual}} = \max_{\mathbf{x}\in\mathbf{R}^n} -f^*(-\mathbf{x}) - (g^{\nu})^*(\mathbf{x}) \quad \text{with} \quad (g^{\nu})_i^*(x_i) = \begin{cases} 0 & \text{if } i \in \mathcal{S}_0 \\ h_i^*(x_i) - \lambda & \text{if } i \in \mathcal{S}_1 \\ \text{conjugate} & \text{if } i \in \mathcal{S}_\bullet \end{cases}$$

1D illustration



Dual Relaxations

Spotlight Result 2

The conjugate admits a generic closed-form expression.

Let $g_i(x) = h_i(x) + \lambda\|x\|_0$, we have

$$g_i^*(x) = \begin{cases} 0 & \text{if } |x| \leq \tau_i \\ h_i^*(x) - \lambda & \text{otherwise} \end{cases}$$

where τ_i is (still) an "easy-to-compute" quantity.

Spotlight Result 2

$\boxed{\textbf{The conjugate admits a generic closed-form expression.}}$

Let $g_i(x) = h_i(x) + \lambda\|x\|_0$, we have

$$g_i^*(x) = \begin{cases} 0 & \text{if } |x| \leq \tau_i \\ h_i^*(x) - \lambda & \text{otherwise} \end{cases}$$

where τ_i is (still) an "easy-to-compute" quantity.

$\downarrow$

Efficient solution methods for the dual relaxation

Proximal gradient

Coordinate descent

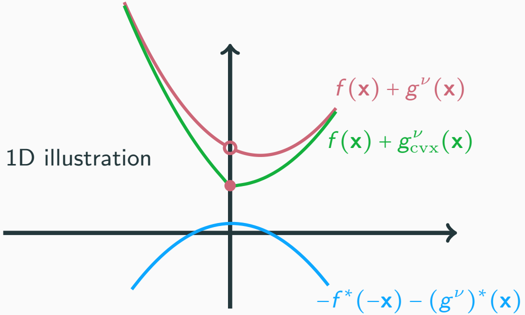
Splitting methods

...

Spotlight Result 3

Convex relaxations are stronger than dual relaxations.

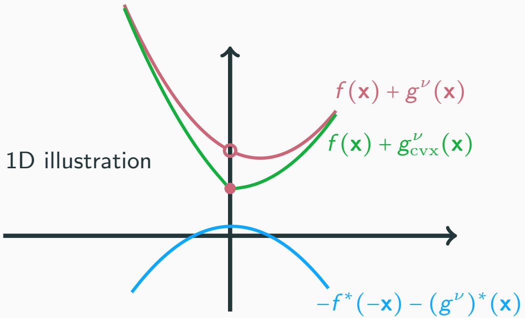
1D illustration



Dual Relaxations

Spotlight Result 3

Convex relaxations are stronger than dual relaxations.



But...

- Evaluating the dual objective always gives a valid lower-bound
- No need to solve the dual relaxation to optimality
- Simple dual link between regions: **simultaneous pruning**

Can we do better?

Can we do better?

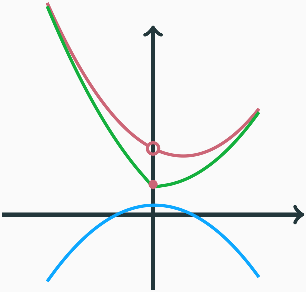
$$p^{\nu} = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + g^{\nu}(\mathbf{x})$$

$$\geq$$

$$p^{\nu}_{\mathrm{cvx}} = \min_{\mathbf{x} \in \mathbf{R}^n} f(\mathbf{x}) + g^{\nu}_{\mathsf{cvx}}(\mathbf{x})$$

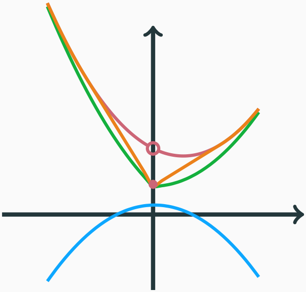
$$\geq$$

$$p^{\nu}_{\mathrm{dual}} = \max_{\mathbf{x} \in \mathbf{R}^m} -f^*(-\mathbf{x}) - (g^{\nu})^*(\mathbf{x})$$



Can we do better?

$$
\begin{aligned}
p^{\nu} &= \min_{\mathbf{x} \in \mathbf{R}^n} \; f(\mathbf{x}) + g^{\nu}(\mathbf{x}) \\
\vee\!| \\
p^{\nu}_{\mathrm{cvx}} &= \min_{\mathbf{x} \in \mathbf{R}^n} \; f(\mathbf{x}) + g^{\nu}_{\mathsf{cvx}}(\mathbf{x}) \\
\vee\!| \\
p^{\nu}_{\mathrm{dual}} &= \max_{\mathbf{x} \in \mathbf{R}^m} \; -f^*(-\mathbf{x}) - (g^{\nu})^*(\mathbf{x})
\end{aligned}
$$



Can we do better?

$$p^\nu = \min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + g^\nu(\mathbf{x})$$

$$\vee\!|$$

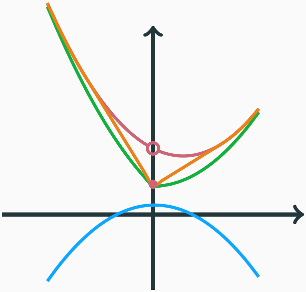
$$p^\nu_{\mathrm{ncvx}} = \min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + g^\nu_{\mathrm{ncvx}}(\mathbf{x})$$

$$\vee\!|$$

$$p^\nu_{\mathrm{cvx}} = \min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + g^\nu_{\mathsf{cvx}}(\mathbf{x})$$

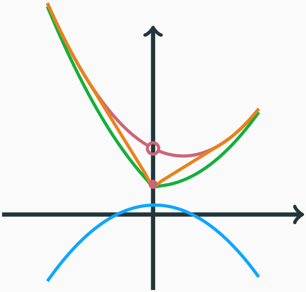
$$\vee\!|$$

$$p^\nu_{\mathrm{dual}} = \max_{\mathbf{x}\in\mathbf{R}^m} -f^*(-\mathbf{x}) - (g^\nu)^*(\mathbf{x})$$



Can we do better?

$$p^\nu = \min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + g^\nu(\mathbf{x})$$
$$\vee\!|$$
$$p^\nu_{\mathrm{ncvx}} = \min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + g^\nu_{\mathrm{ncvx}}(\mathbf{x})$$
$$\vee\!|$$
$$p^\nu_{\mathrm{cvx}} = \min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + g^\nu_{\mathrm{cvx}}(\mathbf{x})$$
$$\vee\!|$$
$$p^\nu_{\mathrm{dual}} = \max_{\mathbf{x}\in\mathbf{R}^m} -f^*(-\mathbf{x}) - (g^\nu)^*(\mathbf{x})$$



Reminder

The relaxation needs to be valid, tight and tractable

Can we do better?

$$p^{\nu} = \min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + g^{\nu}(\mathbf{x})$$

$$\vee\!|$$

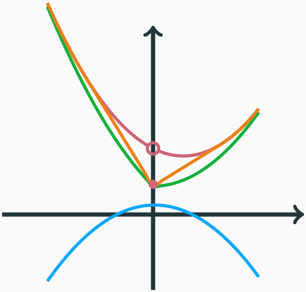
$$p^{\nu}_{\mathrm{ncvx}} = \min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + g^{\nu}_{\mathrm{ncvx}}(\mathbf{x})$$

$$\vee\!|$$

$$p^{\nu}_{\mathrm{cvx}} = \min_{\mathbf{x}\in\mathbf{R}^n} f(\mathbf{x}) + g^{\nu}_{\mathrm{cvx}}(\mathbf{x})$$

$$\vee\!|$$

$$p^{\nu}_{\mathrm{dual}} = \max_{\mathbf{x}\in\mathbf{R}^m} -f^*(-\mathbf{x}) - (g^{\nu})^*(\mathbf{x})$$



Reminder

The relaxation needs to be valid, tight and tractable

1. Validity : $g^{\nu}_{\mathrm{ncvx}} \leq g^{\nu}$
2. Tightness : $g^{\nu}_{\mathrm{ncvx}} \geq g^{\nu}_{\mathrm{cvx}}$
3. Tractability: $f + g^{\nu}_{\mathrm{ncvx}}$ is convex

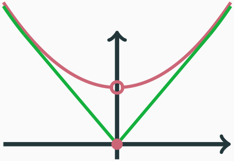
Can we do better?

Least-squares loss and ℓ_2-norm penalty

$$p^\star = \min_{\mathbf{x}\in\mathbf{R}^n} \tfrac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_0 + \tfrac{\gamma}{2}\|\mathbf{x}\|_2^2$$

$$\vee|$$

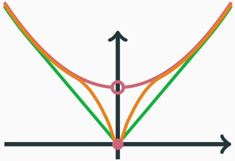
$$p^\star_{\mathrm{cvx}} = \min_{\mathbf{x}\in\mathbf{R}^n} \tfrac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \textsc{Berhu}_{\lambda,\gamma}(\mathbf{x})$$



Can we do better?

Least-squares loss and ℓ_2-norm penalty

$$
\begin{aligned}
p^\star &= \min_{\mathbf{x}\in\mathbf{R}^n} \ \tfrac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_0 + \tfrac{\gamma}{2}\|\mathbf{x}\|_2^2 \\
&\ \vee| \\
p^\star_{\mathrm{ncvx}} &= \min_{\mathbf{x}\in\mathbf{R}^n} \ \tfrac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \mathrm{MCP}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbf{x}) + \tfrac{\gamma}{2}\|\mathbf{x}\|_2^2 \\
&\ \vee| \\
p^\star_{\mathrm{cvx}} &= \min_{\mathbf{x}\in\mathbf{R}^n} \ \tfrac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \mathrm{BERHU}_{\lambda,\gamma}(\mathbf{x})
\end{aligned}
$$



Can we do better?

Least-squares loss and ℓ_2-norm penalty

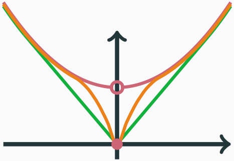
$$p^\star = \min_{\mathbf{x}\in\mathbf{R}^n} \tfrac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_0 + \tfrac{\gamma}{2}\|\mathbf{x}\|_2^2$$

$$\vee\!|$$

$$p^\star_{\mathrm{ncvx}} = \min_{\mathbf{x}\in\mathbf{R}^n} \tfrac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \mathrm{M}\scriptstyle{\mathrm{CP}}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\mathbf{x}) + \tfrac{\gamma}{2}\|\mathbf{x}\|_2^2$$

$$\vee\!|$$

$$p^\star_{\mathrm{cvx}} = \min_{\mathbf{x}\in\mathbf{R}^n} \tfrac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \mathrm{B}\scriptstyle{\mathrm{ERHU}}_{\lambda,\gamma}(\mathbf{x})$$



Sufficient conditions:

- $\boldsymbol{\alpha} = \sqrt{2\lambda\boldsymbol{\beta}}$ $\rightarrow$ validity and tightness ✔
- $\mathbf{A}^{\mathrm{T}}\mathbf{A} + \gamma\mathbf{I} - \mathrm{diag}(\boldsymbol{\beta}) \succeq \mathbf{0}$ $\rightarrow$ tractability ✔

Takeways

- Generalizing BnB ≡ Generic relaxation in a given region
    - Generic convex relaxation
    - Generic dual relaxation
    - Both having their own advantages
- Can we do better?
    - Non-convex relaxation of the penalty
    - Tune parameters to ensure validity, tightness and tractability
    - Case with least-squares loss and ℓ_2-norm penalty
    - We'll see next if it works in practice!

Question time!

We are building a benchmark for sparse problems with

BenchOpt

We are looking for contributors, datasets, solvers, ...

Feel free to join!