

NOML09: DEFLECTED SUBGRADIENT METHODS FOR DUAL FORMULATIONS OF CONVEX QUADRATIC SEPARABLE MIN COST FLOW BOXED PROBLEMS

The Problem and The Approach

We explore some possible dual approaches to solve the Min Cost Flow box constrained convex quadratic separable problem (P) defined by

$$\nu_*(P) = \min_{x \in S_{(P)}} \frac{1}{2} x^\top Q x + q^\top x$$

where $S_{(P)} : Ex = b \quad \cap \quad l \leq x \leq u$

where E , of size (m, n) , is the node-arc incidence matrix of a directed graph and $Q \succcurlyeq 0$ is a diagonal matrix. We choose one of the simplest dual reformulations and approach its iterative solution implementing some subgradient methods; hence we touch upon deflected methods, automatic parameter tuning, and one of the "well known" properties of subgradient methods in dual reformulations.

1. Dual Reformulations

(D1) Dual Flux Constraints Since box constraints are easy, the first option is to dualise only the flux constraints:

$$\nu_*(P) = \min_{x \in S_{(D1)}} \max_{\mu} L(x, \mu) \geq \max_{\mu} \min_{x \in S_{(D1)}} L(x, \mu) = \nu^*(D1)$$

where $L(x, \mu) = \frac{1}{2} x^\top Q x + q^\top x + \mu^\top (Ex - b)$

and $S_{(D1)} : l \leq x \leq u$

Since the feasible space is a box and the target function is separable and convex, to find

$$L(\mu) = \min_{x \in S_{(D1)}} L(x, \mu) = L(\bar{x}, \mu)$$

when Q is positive definite we can solve for x in

$$\nabla_x L(\bar{x}, \mu) = Q\bar{x} + q + E^\top \mu = 0$$

and then project to the nearest side, component by component:

$$x = \max.(l, \min.(u, -Q^{-1}(q + E^\top \mu)))$$

For the coordinates corresponding to $\ker Q$, we should look for

$$\min_x (q + E^\top \mu)_j x_j \implies \begin{cases} (q + E^\top \mu)_j > 0 \implies x_j = l_j \\ (q + E^\top \mu)_j < 0 \implies x_j = u_j \\ x_j \in [l_j, u_j] \quad \text{otherwise} \end{cases}$$

The specific value of x_j that we choose corresponds to a choice of the super-gradient $\partial_\mu L(\mu) = Ex(\mu) - b$, where $x(\mu)$ can be considered a multivalued function, therefore $L(\mu)$ is "piecewise differentiable" but still easy to calculate.

(D2) Dual Flux and Box Constraints We start from the Lagrangian

$$L(x, \mu, \lambda) = \frac{1}{2} x^\top Q x + q^\top x + \mu^\top (Ex - b) + \lambda_u^\top (x - u) + \lambda_l^\top (l - x)$$

with $\lambda. \geq 0$

Since the linear coefficients of x will determine the position of the minimum of L for fixed dual parameters, a possible substitution to make clear the relevant degrees of freedom is

$$\begin{aligned} L(x, \mu, \gamma, \lambda_u) &= \frac{1}{2} x^\top Q x + (\gamma + q)^\top x - \mu^\top b + (E^\top \mu - \gamma)^\top l + \lambda_u^\top (l - u) \\ &\quad \text{with} \quad \lambda_u. \geq 0 \quad \& \quad \lambda_u + E^\top \mu - \gamma. \geq 0 \\ \iff \quad \gamma &= \Delta \lambda + E^\top \mu \quad \text{where} \quad \Delta \lambda = \lambda_u - \lambda_l \end{aligned}$$

In the coordinate subspace defined by $\ker Q$ the Lagrangian is unbounded below unless $\nabla_x|_{\ker Q} L(x, \mu, \gamma, \lambda_u) = 0$. Since Q is diagonal, we can decompose the coordinate space in the kernel and the co-image by a simple partition of the coordinates, corresponding in obvious notations to $Q_0, x_0 \dots$ and Q_1, x_1, \dots . Hence, *for the interesting (finite) minima of L varying x* , we have

$$\begin{aligned} Q\bar{x}(\gamma) + \gamma + q &= 0 \implies \\ L(\mu, \gamma, \lambda_u) &= L(\bar{x}(\gamma), \mu, \gamma, \lambda_u) = \min_x L(x, \mu, \gamma, \lambda_u) = \dots \\ \dots &= -\frac{1}{2}(\gamma + q)_1^\top Q_1^{-1}(\gamma + q)_1 - \mu^\top b + (E^\top \mu - \gamma)^\top l + \lambda_u^\top (l - u) = \dots \\ \dots &= -\frac{1}{2}(\gamma + q)_1^\top Q_1^{-1}(\gamma + q)_1 - \mu^\top b - (\gamma + q)^\top l + (E^\top \mu + q)^\top l + \lambda_u^\top (l - u) \end{aligned}$$

Note that

$$Q\bar{x} + \gamma + q = 0 \iff \gamma_0 + q_0 = 0 \quad \& \quad Q_1 \bar{x}_1 + \gamma_1 + q_1 = 0$$

With the substitution $\tilde{\gamma} = \gamma + q$, our objective function is written

$$\nu_*(D2) = \max_{\tilde{\gamma}, \mu, \lambda_u} L(\tilde{\gamma}, \mu, \lambda_u) \quad \text{beholding}$$

$$\lambda_u. \geq \max.(0, \tilde{\gamma} - q - E^\top \mu) \quad \& \quad \tilde{\gamma}_0 = 0$$

$$\text{where } L(\tilde{\gamma}, \mu, \lambda_u) = -\frac{1}{2} \tilde{\gamma}_1^\top Q_1^{-1} \tilde{\gamma}_1 - \mu^\top b - \tilde{\gamma}^\top l + (E^\top \mu + q)^\top l + \lambda_u^\top (l - u)$$

We are looking for the maximum, we know that $l - u. \geq 0$ and we have the constraint $\lambda_u. \geq \max.(0, \tilde{\gamma} - q - E^\top \mu)$, so, in the last formula for the Lagrangian, the last term $\lambda_u^\top (l - u)$ can be restricted to an equality, leading to

$$\max_{\tilde{\gamma}, \mu, \lambda_u} L(\tilde{\gamma}, \mu, \lambda_u) = \max_{\tilde{\gamma}, \mu} L(\tilde{\gamma}, \mu) \quad \text{where}$$

$$\begin{aligned} L(\tilde{\gamma}, \mu) &= -\frac{1}{2} \tilde{\gamma}_1^\top Q_1^{-1} \tilde{\gamma}_1 - \mu^\top b - \tilde{\gamma}^\top l + (E^\top \mu + q)^\top l + \max.(0, \tilde{\gamma} - q - E^\top \mu)^\top (l - u) = \dots \\ &\dots = -\frac{1}{2} \tilde{\gamma}_1^\top Q_1^{-1} \tilde{\gamma}_1 - \mu^\top b - (\tilde{\gamma} - q - E^\top \mu)^\top (l + (u - l). * \theta.(\tilde{\gamma} - q - E^\top \mu)) \end{aligned}$$

where $a\theta(a) = \max(0, a)$. Since Q_1 is diagonal, we have

$$L(\tilde{\gamma}, \mu) = \sum_i L_i(\tilde{\gamma}, \mu) - \mu^\top b \quad \text{where}$$

$$L_i(\tilde{\gamma}, \mu) = -\frac{1}{2} \tilde{\gamma}_i Q_{ii}^{-1} \tilde{\gamma}_i - (\tilde{\gamma} - q - E^\top \mu)_i \begin{cases} l_i & \text{when } (\tilde{\gamma} - q - E^\top \mu)_i \leq 0 \\ u_i & \text{when } (\tilde{\gamma} - q - E^\top \mu)_i \geq 0 \end{cases}$$

which is an unconstrained, continuous and piecewise differentiable formulation of the problem. Recall that $(\tilde{\gamma} - q - E^\top \mu) = \Delta \lambda$. The optimality conditions can be expressed in terms of the supergradient: we denote with p_i the coordinate l_i if $\Delta \lambda_i < 0$, u_i if $\Delta \lambda_i > 0$ and a value in $[l_i, u_i]$ otherwise. Then, when the supergradient conditions

$$Qp + \tilde{\gamma} = 0 \quad \& \quad Ep - b = 0$$

are met, we obtain

$$L = -\mu^\top b + \frac{1}{2} p^\top Qp + (E^\top \mu + q)^\top p = \frac{1}{2} p^\top Qp + q^\top p$$

Duality is there.

(D3) Flux Constraints in Parametric Form, Dual Box Constraints

From the incidence matrix of a connected graph we can easily extract a submatrix of maximum degree by choosing a spanning tree; then, chosen a root, its inverse is described in each node by the subtree rooted in such node¹.

TBW

¹Given a root, the sparsest inverse with such node excluded is found by choosing the BFS tree, so the submatrix of maximum rank with sparsest inverse can be found by comparing the BFS trees of each node

Fancier Duals Among the many other dual reformulations of the problem, the most appealing could probably be the smoothing ones, which unlock the use of faster descent methods [3]. We are not covering them since this work is explicitly required to deal with subgradient methods.

2. Algorithms and Implementation

We choose the dual reformulation (D1) as the main testing bed for the many iterative solvers we implement. Since an exact line search is possible, our main threads of exploration can be divided into the ones using a line search and the other ones using proper subgradient methods.

Conjugate Subgradient with Exact Line Search We apply Polak-Ribière direction update endowing the solver with a projected conjugate gradient algorithm to find the minimum norm subgradient {3}. The exact line search {4} exploits a priority queue to pop in the right order the intersection points where the hessian of the dual Lagrangian is changing. The sorting criterion checks for consistency and imposes an ordering coherent with the geometry up to a given ϵ , with the goal of curing the inconsistencies deriving from floating point errors. Nonetheless, the implementation needs further revisions to tackle straightly with a strongly singular Q (problem which can be cured with the simplest of the smoothing methods, that is reaching iteratively the singular Q).

Subgradient Methods We experiment with many different subgradient and deflected subgradient steps. Parameter tuning is a game changer here and the performance of some algorithms is extremely sensible to parameters {1, 2}. Hence all the subgradient steps are implemented so that they can be controlled by an external tuner. At the moment the only parameter search which is implemented is the standard Nelder Mead; further work is required to adapt it to the constrained case, though.

Implementation The software is implemented as a package in Julia. Companion to this report a Jupyter notebook, illustrating the use of the package with some examples.

Bibliography

- [1] Frangioni A., Gendron B., Gorgone E., “On the Computational Efficiency of Subgradient Methods: a Case Study with Lagrangian Bounds”, *Mathematical Programming Computation*, **9:4** (2017), 573–604.
- [2] Gustavsson E., Patriksson M., Strömberg A., “Primal convergence from dual subgradient methods for convex optimization”, *Mathematical Programming*, **150** (2014).
- [3] Nesterov Y., “Universal gradient methods for convex optimization problems”, *Mathematical Programming*, **152** (2015), 381–404.
- [4] Boyd S., “Subgradient Methods”, *Notes for EE364b, Stanford University*, 2014.

- [5] Stetsyuk P., “Accelerated subgradient method with Polyak’s step”, *Workshop*, 2018.

Lapo Toloni, e-mail: l.toloni@studenti.unipi.it

Michele Miccinesi, e-mail: m.miccinesi@studenti.unipi.it

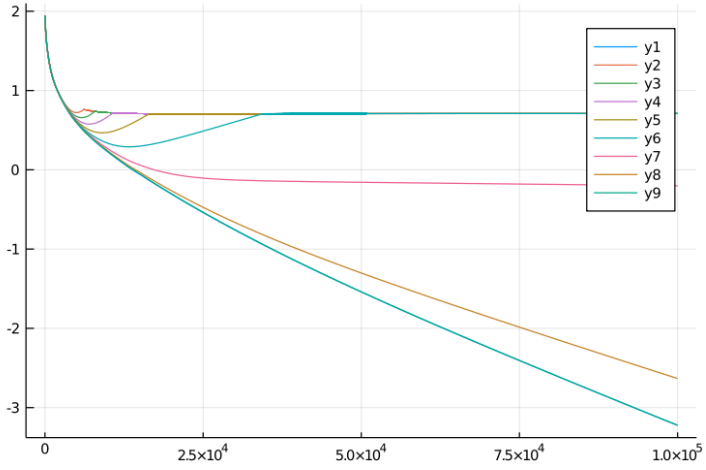


Figure 1. $\|\partial L\|$ Log10-scale objective minimization with varying parameters of the Nesterov Momentum algorithm. Here the parameters are encompassing a relative variation of only 10^{-8} , but the algorithm dynamic is completely changing

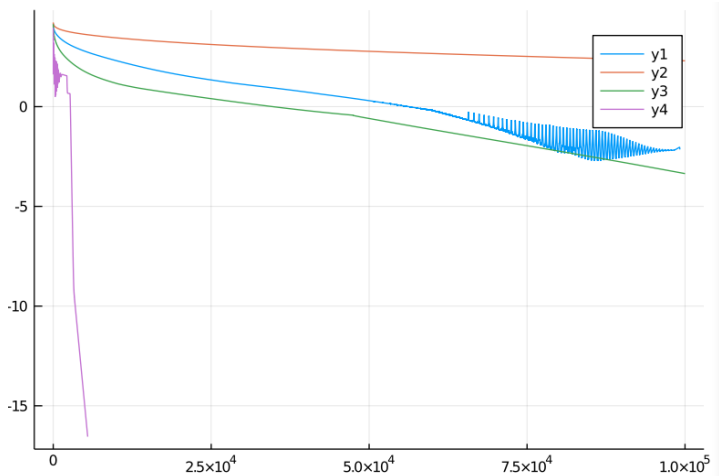


Figure 2. $\|\partial L\|$ Log10-scale objective minimization with varying parameters of the RMSProp algorithm. Usually, varying the parameters of such algorithms, it is possible to reach a "stationary point" for the dynamics; sometimes, with a stroke of luck, as in this case, moving a bit further may lead to a point where the dynamics is initially less stable, afterward incredibly faster

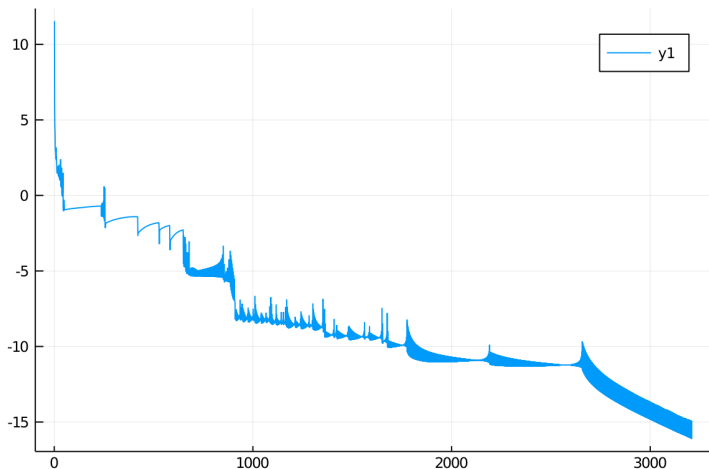


Figure 3. $\|Ex - b\|^2$ Log10-scale objective minimization typical dynamics of projected conjugate gradient with long search + local search, here on a 1000×2000 incidence matrix E .

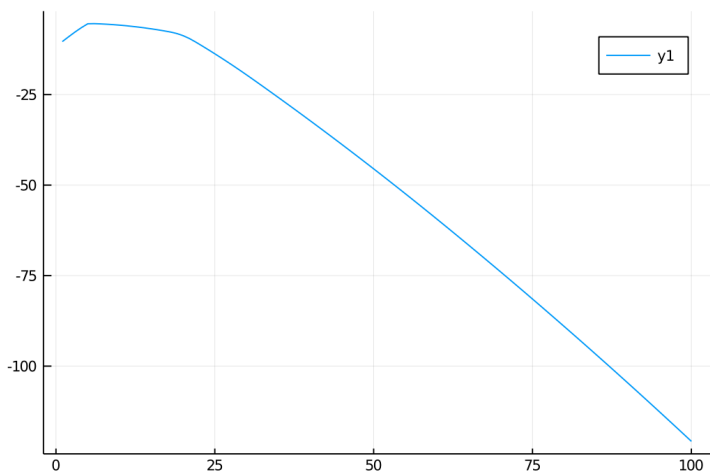


Figure 4. $L(\mu + \alpha d)$ Lagrangian dual (D1) restricted to the direction of line search when Q is singular; here the primal x corresponds to the supergradient we choose on the top vertex.