

Notes - Interior Point Polynomial Time Methods for Linear Programming, Conic Quadratic Programming, and Semidefinite Programming

Vitoria Guardieiro

December 2021

1 Motivation

Generic convex problems, under mild computability and boundedness assumptions, are polynomially solvable, e.g., by the ellipsoid method. It turns out that both the strong (universality) and the weak (poor practical performance) features of the ellipsoid method have the same origin — the fact that the method deals with black box–represented problems. When starting the solution process, the method has no idea what problem it is working with. All it knows is how many design variables are in the problem and how large and thin its feasible set can be.

In practice, there are no entities like “a convex objective f ” and “a convex feasible set \mathcal{X} ”, with all our a priori knowledge of f and \mathcal{X} being expressed by the word “convex”. Normally we deal with instances of a known-in-advance generic convex program, like LP, CQP, or SDP, with known data and therefore we possess from the very beginning complete global information on the problem to be processed. What we would like to have is an optimization technique able to utilize efficiently our global knowledge of the instance and thus able to generate a solution much faster than a nearly blind algorithm like the ellipsoid method. The major event in the recent history of convex optimization, called sometimes interior point revolution, was the invention of such smart techniques.

2 The interior penalty scheme

Consider a constrained convex optimization program. One can w.l.o.g. make its objective linear, moving, if necessary, the actual objective to the list of constraints. Thus, let the problem be

$$\min_x \{c^T x : x \in \mathcal{X} \in \mathbb{R}^n\}, \quad (\text{C})$$

where \mathcal{X} is a closed convex set, which we assume to possess a nonempty interior.

Let us choose a barrier (also called an interior penalty function) $F(x)$ for the feasible set \mathcal{X} , which is a function that is well defined, smooth, strongly convex on the interior of \mathcal{X} and that blows up as a point from $\text{int}\mathcal{X}$ approaches a boundary point of \mathcal{X} :

$$x_i \in \text{int}\mathcal{X}, x \equiv \lim_{i \rightarrow \infty} x_i \in \partial\mathcal{X} \implies \lim_{i \rightarrow \infty} F(x_i) = \infty.$$

We next consider the one-parametric family of functions generated by our objective and the barrier:

$$F_t(x) = tc^T x + F(x) : \text{int}\mathcal{X} \rightarrow \mathbb{R}$$

Here t is the penalty parameter, which is assumed to be nonnegative.

Under mild regularity assumptions (e.g., in the case of bounded \mathcal{X} , which we assume from now on), we have that:

- Every function $F_t(\cdot)$ attains its minimum over the interior of \mathcal{X} , the minimizer $x_*(t)$ being unique.
- The central path $x_*(t)$ is a smooth curve, and all its limiting, $t \rightarrow \infty$, points belong to the set of optimal solutions of (C).

To minimize $F_t(\cdot)$ for large t is the same as minimizing the function $f_\rho(x) = f(x) + \rho F(x)$ for small $\rho = 1/t$. When ρ is small, the function f_ρ is very close to f everywhere in \mathcal{X} except a narrow stripe along the boundary of \mathcal{X} , the stripe becoming thinner and thinner as $\rho \rightarrow 0$. Therefore we have good reasons to believe that the minimizer of F_t for large t (i.e., the minimizer of f_ρ for small ρ) must be close to the set of minimizers of f on \mathcal{X} . The central path $x_*(t)$ is a kind of Ariadne's thread that leads to the solution set of (C).

Given a value $t \geq 0$ of the penalty parameter, to reach the point $x_*(t)$ on this path is the same as minimizing a smooth strongly convex function $F_t(\cdot)$ which attains its minimum at an interior point of \mathcal{X} . The latter problem is nearly unconstrained, up to the fact that its objective is not everywhere defined. However, we can easily adapt the methods of unconstrained minimization, including the Newton one, to handle nearly unconstrained problems. We see that constrained convex optimization in a sense can be reduced to the easy unconstrained one.

Assume that we know in advance the minimizer of $F_0 \equiv F$, i.e., the point $x_*(0)$. Thus, we know where the central path starts. Now let us follow this path: at i th step, standing at a point x_i close enough to some point $x_*(t_i)$ of the path, we

- first, increase a bit the current value t_i of the penalty parameter, thus getting a new target point $x_*(t_{i+1})$ on the path; and
- second, approach our new target point $x_*(t_{i+1})$ by running, say, the Newton method on the function $F_{t_{i+1}}$ and starting the method at our current iterate x_i , until a new iterate x_{i+1} close enough to $x_*(t_{i+1})$ is generated.

Iterating this updating and strengthening appropriately our closeness requirements as the process goes on, we approach the optimal set along the central path. If all our requirements of “close enough” and “not too rapidly” are properly controlled, we may ensure x_i to be in the domain of quadratic convergence of the Newton method as applied to $F_{t_{i+1}}$, which means that it will take a quite small number of steps to recover closeness to our new target point.

However, as far as complexity is concerned, for nearly all black box-represented classes of unconstrained convex optimization problems, there is no such phenomenon as local quadratic convergence, the Newton method (which uses the second order derivatives) has no advantages over the methods that use only first order derivatives.

What prevents this scheme from yielding a polynomial time method is not the structure of the scheme but the complete freedom it allows for some of its elements. After some order is added, the scheme becomes a polynomial time one. It turns out that we can do that for all interesting well-structured generic convex programs, in particular, for linear, conic quadratic, and semidefinite programming.

3 Interior point methods for linear programming, conic quadratic programming, and semidefinite programming: Building blocks

3.1 Canonical cones and canonical barriers

We will be interested in a generic conic problem

$$\min_x \{c^T x : Ax - B \in \mathbf{K}\} \quad (\text{CP})$$

associated with a cone \mathbf{K} given as a direct product of m basic cones, each being either a Lorentz (\mathbf{L}) or a semidefinite cone (\mathbf{S}_+):

$$\mathbf{K} = \mathbf{S}_+^{k_1} \times \cdots \times \mathbf{S}_+^{k_p} \times \mathbf{L}^{k_{p+1}} \times \cdots \times \mathbf{L}^{k_m} \quad (\text{Cone})$$

The generic problem in question covers LP (no Lorentz cones in the right-hand side, and all semidefinite cones are of dimension 1), CQP (no semidefinite cones), and SDP (no Lorentz cones).

We equip the semidefinite and the Lorentz cones with canonical barriers:

- Canonical barrier for a semidefinite cone \mathbf{S}_+^k :

$$S_k(X) = -\ln \text{Det}(X) : \text{int}\mathbf{S}_+^k \rightarrow \mathbb{R};$$

the parameter of this barrier, by definition, is $\theta(S_k) = k$.

- Canonical barrier for a Lorentz cone \mathbf{L}^k :

$$L_k(x) = -\ln(x_k^2 - x_1^2 - \cdots - x_{k-1}^2);$$

with parameter $\theta(L_k) = 2$.

- Canonical barrier for the cone \mathbf{K} given by (Cone):

$$K(X) = S_{k_1}(X_1) + \cdots + S_{k_p}(X_p) + L_{k_{p+1}}(X_{p+1}) + \cdots + L_{k_m}(X_m),$$

$$X_i \in \begin{cases} \text{int}\mathbf{S}_+^{k_i}, & i \leq p \\ \text{int}\mathbf{L}^{k_i}, & p < i \leq m. \end{cases}$$

The parameter of the barrier $K(\cdot)$ is the sum of parameters of the basic barriers involved:

$$\theta(K) = \theta(S_{k_1}) + \cdots + \theta(S_{k_p}) + \theta(L_{k_{p+1}}) + \cdots + \theta(L_{k_m}) = \sum_{i=1}^p k_i + 2(m - p).$$

We can see that our basic barriers (and therefore their direct sum $K(\cdot)$) are indeed barriers for the corresponding cones: they are C^∞ -smooth on the interiors of their domains, blow up to ∞ along every sequence of points from these interiors converging to a boundary point of the corresponding domain, and are strongly convex.

3.2 Elementary properties of canonical barriers

Let us establish a number of simple and useful properties of canonical barriers.

Proposition 3.1. *A cononical barrier F (F can be S_k , L_k , or the direct sum K of several copies of these elementary barriers) possesses the following properties:*

- (i) *F is logarithmically homogeneous with the parameter of logarithmic homogeneity equal to $-\theta(F)$, i.e., the following identity holds:*

$$t > 0, x \in \text{Dom}F \implies F(tx) = F(x) - \theta(F) \ln t.$$

- (ii) *The following two equalities hold identically in $x \in \text{Dom}F$:*

- (a) $\langle \nabla F(x), x \rangle = -\theta(F)$,
(b) $[\nabla^2 F(x)]x = -\nabla F(x)$.

- (iii) *The l^{th} differential $D^l F(x)$ of F , $l \geq 1$, is homogeneous of degree $-l$ in $x \in \text{Dom}F$:*

$$\forall (x \in \text{Dom}F, t > 0, h_1, \dots, h_l) :$$

$$D^l F(tx)[h_1, \dots, h_l] \equiv \left. \frac{\partial^l F(tx + s_1 h_1 + \dots + s_l h_l)}{\partial s_1 \partial s_2 \dots \partial s_l} \right|_{s_1 = \dots = s_l = 0} = t^{-l} D^l F(x)[h_1, \dots, h_l]. \quad (1)$$

Proof. (i) Both S_k and L_k are clearly are logarithmically homogeneous with the parameters of logarithmic homogeneity $-\theta(S_k)$, $-\theta(L_k)$, respectively. Also, the property of logarithmic homogeneity is stable with respect to taking direct sums of functions, so the direct sum K is also logarithmically homogeneous.

- (ii) To get (ii)(a), it suffices to differentiate the identity

$$F(tx) = F(x) - \theta(F) \ln t$$

in t at $t = 1$,

$$\langle \nabla F(tx), x \rangle = \frac{d}{dt} F(tx) = -\theta(F) t^{-1}$$

and it remains to set $t = 1$ in the concluding identity.

Similarly, to get (ii)(b), it suffices to differentiate the identity

$$\langle \nabla F(x + th), x + th \rangle = -\theta(F)$$

(which is just (ii)(a)) in t at $t = 0$, thus arriving at

$$\langle [\nabla^2 F(x)]h, x \rangle + \langle \nabla F(x), h \rangle = 0$$

since $\langle [\nabla^2 F(x)]h, x \rangle = \langle [\nabla^2 F(x)]x, h \rangle$ and since the resulting equality

$$\langle [\nabla^2 F(x)]x, h \rangle + \langle \nabla F(x), h \rangle = 0$$

holds true identically in h , we come to $[\nabla^2 F(x)]x = -\nabla F(x)$.

- (iii) Differentiating l times the identity

$$F(tx) = F(x) - \theta \ln t$$

in x , we get

$$t^l D^l F(tx)[h_1, \dots, h_l] = D^l F(x)[h_1, \dots, h_l].$$

□

An especially nice specific feature of canonical barriers is their self-duality:

Proposition 3.2. *A canonical barrier F possesses the following property: for every $x \in \text{Dom}F$, $-\nabla F(x)$ belongs to $\text{Dom}F$ as well, and the mapping $x \mapsto -\nabla F(x) : \text{Dom}F \rightarrow \text{Dom}F$ is self-inverse,*

$$-\nabla F(-\nabla F(x)) = x \quad \forall x \in \text{Dom}, \quad (2)$$

and is homogeneous of degree -1 :

$$t > 0, x \in \text{intdom}F \implies -\nabla F(tx) = -t^{-1}\nabla F(x). \quad (3)$$

4 Primal-dual pair of problems and the primal-dual central path

4.1 Problem(s)

Since \mathbf{K} is a direct product of self-dual cones, this dual is a conic problem on the same cone \mathbf{K} . As we remember, the primal-dual pair associated with (CP) is

$$\min_x \{c^T x : \mathcal{A}x - B \in \mathbf{K}\} \quad (\text{CP})$$

$$\max_S \{\langle B, S \rangle_E : \mathcal{A} * S = c, S \in \mathbf{K}\} \quad (\text{CD})$$

Assuming that $\text{Null}(\mathcal{A}) = \{0\}$, we can write down our primal-dual pair in a symmetric geometric form:

$$\min_X \{\langle C, X \rangle_E : X \in (\mathcal{L} - B) \cap \mathbf{K}\} \quad (\text{P})$$

$$\max_S \{\langle B, S \rangle_E : S \in (\mathcal{L}^\perp + B) \cap \mathbf{K}\} \quad (\text{D})$$

where \mathcal{L} is a linear subspace in E (the image space of the linear mapping $x \mapsto \mathcal{A}x$), \mathcal{L}^\perp is the orthogonal complement to \mathcal{L} in E , and $C \in E$ satisfies $\mathcal{A} * C = c$, i.e., $\langle C, \mathcal{A}x \rangle_E \equiv c^T x$.

To simplify things, from now on we assume that both problems (CP) and (CD) are strictly feasible. In terms of (P) and (D) this assumption means that both the primal feasible plane $\mathcal{L} - B$ and the dual feasible plane $\mathcal{L}^\perp + B$ intersect the interior of the cone \mathbf{K} .

4.2 Central path(s)

The canonical barrier K of \mathbf{K} induces a barrier for the feasible set $\mathcal{X} = \{x \mid \mathcal{A}x - B \in \mathbf{K}\}$ of the problem (CP) written down in the form of (C), i.e., as

$$\min_x \{c^T x : x \in \mathcal{X}\};$$

this barrier is

$$\hat{K}(x) = K(\mathcal{A}x - B) : \text{int}X \rightarrow \mathbb{R}$$

and is indeed a barrier. Now we can apply the interior penalty scheme to trace the central path $x_*(t)$ associated with this barrier.

It can be derived from the primal-dual strict feasibility that the central path is well defined, i.e., that the minimizer of

$$\hat{K}_t(x) = tc^T x + \hat{K}(x)$$

on $\text{int}X$ exists for every $t \geq 0$ and is unique.

It is highly instructive to pass from the central path $x_*(t)$ in the space of design variables to its image in E :

$$X_*(t) = \mathcal{A}x_*(t) - B.$$

The resulting curve is called the *primal central path* of the primal-dual pair (P), (D); by its origin, it is a curve comprising strictly feasible solutions of (P).

Observation 4.1. *A point $X_*(t)$ of the primal central path is the minimizer of the aggregate*

$$P_t(X) = t\langle C, X \rangle_E + K(X)$$

on the set $(\mathcal{L} - B) \cap \text{int}\mathbf{K}$ of strictly solutions of (P).

Observation 4.2. *A point $X_*(t)$ of the primal central path is exactly the strictly feasible solution X to (P) such that the vector $tC + \nabla K(X) \in E$ is orthogonal to \mathcal{L} (i.e., belongs to \mathcal{L}^\perp).*

The dual problem (D) also possesses the central path, now called the *dual central path* $S_*(t), t \geq 0$, of the primal-dual pair (P), (D). Similar to 4.1 and 4.2, the dual central path can be characterized as follows:

Observation 4.3. *A point $S_*(t), t \geq 0$, of the dual central path is the unique minimizer of the aggregate*

$$D_t(S) = -t\langle B, S \rangle_E + K(S)$$

on the set of strictly feasible solutions of (D). $S_(t)$ is exactly the strictly feasible solution S to (D) such that the vector $-tB + \nabla F(S)$ is orthogonal to \mathcal{L}^\perp (i.e., belongs to \mathcal{L}).*

From Proposition 3.2 we can derive a connection between the primal and the dual central paths:

Theorem 4.4. *For $t > 0$, the primal and the dual central paths $X_*(t), S_*(t)$ of a (strictly feasible) primal-dual pair (P), (D) are linked by the relations*

$$\begin{aligned} S_*(t) &= -t^{-1}\nabla K(X_*(t)), \\ X_*(t) &= -t^{-1}\nabla K(S_*(t)). \end{aligned}$$

Proof. By Observation 4.2, the vector $tC + \nabla K(X_*(t))$ belongs to \mathcal{L}^\perp , so the vector $S = -t^{-1}\nabla K(X_*(t))$ belongs to the dual feasible plane $\mathcal{L}^\perp + C$. On the other hand, the vector $-\nabla K(X_*(t))$ belongs to $\text{Dom}K$, i.e., to the interior of \mathbf{K} . Since \mathbf{K} is a cone and $t > 0$, the vector $S = -t^{-1}\nabla K(X_*(t))$ belongs to the interior of \mathbf{K} as well. Thus, S is a strictly feasible solution of (D). Now let us compute the gradient of the aggregate D_t at the point S :

$$\begin{aligned} \nabla D_t(S) &= -tB + \nabla K(-t^{-1}\nabla K(X_*(t))) \\ &= -tB + t\nabla K(-\nabla K(X_*(t))) \\ &= -tB - tX_*(t) \\ &= -t(B + X_*(t)) \\ &\in \mathcal{L}. \end{aligned}$$

Thus, S is strictly feasible for (D) and $\nabla D_t(S) \in \mathcal{L}$. But by Observation 4.3 these properties characterize $S_*(t)$; thus, $S_*(t) = S \equiv -t^{-1}\nabla K(X_*(t))$. In view of Proposition 3.2, this implies that $X_*(t) = -t^{-1}\nabla K(S_*(t))$. Another way to get the latter relation from the relation $S_*(t) = -t^{-1}\nabla K(X_*(t))$ is just to refer to the primal-dual symmetry. \square

Theorem 4.5. *Let (P) , (D) be a strictly feasible primal-dual pair.*

For every $t > 0$, there exists a unique strictly feasible solution X of (P) such that $-t^{-1}\nabla K(X)$ is a feasible solution to (D) , and this solution X is exactly $X_(t)$.*

Similarly, for every $t > 0$, there exists a unique strictly feasible solution of (D) such that $-t^{-1}\nabla K(S)$ is a feasible solution of (P) , and this solution S is exactly $S_(t)$.*

Proof. By primal-dual symmetry, it suffices to prove the first claim. We know from the previous theorem that $X = X_*(t)$ is a strictly solution of (P) such that $-t^{-1}\nabla K(X)$ is feasible for (D) . All we need to prove is that $X_*(t)$ is the only point with these properties, which is immediate: if X is a strictly feasible solution of (P) such that $-t^{-1}\nabla K(X)$ is dual feasible, then $-t^{-1}\nabla K(X) \in \mathcal{L}^\perp + C$, whence $\nabla K(X) \in \mathcal{L}^\perp - tC$, or, which is the same, $\nabla P_t(X) = tC + \nabla K(X) \in \mathcal{L}^\perp$. We already know from Observation 4.2 that the latter property, together with the strict primal feasibility, is characteristic for $X_*(t)$. \square

4.2.1 Characterization of the central path

As we have seen, the primal and dual central paths are intrinsically linked one to another, and it makes sense to treat them as a unique entity—the primal-dual central path of the primal-dual pair (P) , (D) . From now on we refer to the primal-dual central path simply as the *central path*.

The points $(X_*(t), S_*(t))$ fo the central path possess the following properties:

- **CP₁ (primal feasibility):** The point $X_*(t)$ is strictly primal feasible.
- **CP₂ (dual feasibility):** The point $S_*(t)$ is dual feasible.
- **CP₃ (augmented complementary slackness):** The points $X_*(t)$ and $S_*(t)$ are linked by the relation

$$S_*(t) = -t^{-1}\nabla K(X_*(t)) \iff X_*(t) = -t^{-1}\nabla K(S_*(t))$$

Properties CP₁, CP₂, CP₃ fully characterize the central path: if two points X, S possess these properties with respect to some $t > 0$, then $X = X_*(t)$ and $S = S_*(t)$ (Theorem 4.5).

4.2.2 Duality gap along the central path

Proposition 4.6. *Under assumption of primal-dual strict feasibility, the duality gap along the central path is inverse proportional to the penalty parameter, and the proportionality coefficient is the parameter $\theta(K)$ of the canonical barrier K :*

$$\text{DualityGap}(X_*(T), S_*(t)) = \frac{\theta(K)}{t}.$$

In particular, both $X_(t)$ and $S_*(t)$ are strictly feasible $(\theta(K)/t)$ -approximate solutions to thei respective problems:*

$$\begin{aligned} \langle C, X_*(t) \rangle_E - \text{Opt}(P) &\leq \frac{\theta(K)}{t}, \\ \text{Opt}(D) - \langle B, S_*(t) \rangle_E &\leq \frac{\theta(K)}{t}. \end{aligned}$$

Proof.

$$\begin{aligned}
\text{DualityGap}(X_*(t), S_*(t)) &= \langle S_*(t), X_*(t) \rangle_E \\
&= \langle -t^{-1} \nabla K(X_*(t)), X_*(t) \rangle_E \\
&= t^{-1} \theta(K)
\end{aligned}$$

□

4.2.3 Distance to the central path

The conclusion so far is that our life when moving along the central path would be just fine (at the very least, we would know how good the solutions we already have are). But *how close should we be to the path (and in what sense close) for our life to be as nice as if we were exactly on the path?*

Our canonical barrier $K(\cdot)$ is a strongly convex smooth function on $\text{int}\mathbf{K}$; in particular, its Hessian matrix $\nabla^2 K(Y)$, taken at a point $Y \in \text{int}\mathbf{K}$, is positive definite. We can use the inverse of this matrix to measure the distances between points of E , thus arriving at the norm

$$\|H\|_Y^* = \sqrt{\langle [\nabla^2 K(Y)]^{-1} H, H \rangle_E}.$$

It turns out that a good measure of proximity of a strictly feasible primal-dual pair $Z = (X, S)$ to a point $Z_*(t) = (X_*(t), S_*(t))$ from the primal-dual central path is the quantity

$$\text{dist}(Z, Z_*(t)) \equiv \|tS + \nabla K(X)\|_X^* \equiv \sqrt{\langle [\nabla^2 K(X)]^{-1} (tS + \nabla K(X)), tS + \nabla K(X) \rangle_E}$$

Proposition 4.7. *If $Z = (X, S)$ is a pair of primal-dual strictly feasible solutions to (P), (D) such that*

$$\text{dist}(Z, Z_*(t)) \leq 1$$

then Z is essentially as good as $Z_(t)$, namely, the duality gap as (X, S) is essentially as small as at the point $Z_*(t)$:*

$$\text{DualityGap}(X, S) = \langle S, X \rangle_E \leq 2 \text{DualityGap}(Z_*(t)) = \frac{2\theta(K)}{t}.$$

It follows from this that for our purposes, it is essentially the same to move along the primal-dual central path or to trace this path, staying in its time-space neighborhood

$$\mathcal{N}_\kappa = \{(t, X, S) | X \in \mathcal{L} - B, S \in \mathcal{L}^\perp + C, t > 0, \text{dist}((X, S), (X_*(t), S_*(t))) \leq \kappa\}$$

with certain $\kappa \leq 1$.

5 Tracing the central path

5.1 Path-following scheme

Assume we are solving a strictly feasible primal-dual pair of problems (P), (D) and intend to trace the associated central path. For this purpose, we just need a mechanism for updating a current iterate $(\bar{t}, \bar{X}, \bar{S})$ such that $\bar{t} < 0$, \bar{X} is strictly primal feasible, \bar{S} is strictly dual feasible, and (\bar{X}, \bar{S}) is a good approximation of the point $Z_*(\bar{t}) = (X_*(\bar{t}), S_*(\bar{t}))$ on the central path, into a new iterate (t_+, X_+, S_+) with similar properties and a larger value $t_+ > \bar{t}$ of the penalty parameter. Given such an updating mechanism and iterating it, we indeed shall trace the central path.

How could we construct the required updating? Recalling the description of the central path, we rephrase our question as:

Given a triple $(\bar{t}, \bar{X}, \bar{S})$ that satisfies the relations

$$X \in \mathcal{L} - B, S \in \mathcal{L}^\perp + C \quad (4)$$

and approximately satisfies the system of nonlinear equations

$$G_t(X, S) \equiv S + t^{-1} \nabla K(X) = 0, \quad (5)$$

update it into a new triple (t_+, X_+, S_+) with the same properties and such that $t_+ > \bar{t}$.

The most natural way, from the viewpoint of computational mathematics, to achieve our target is as follows:

1. Choose somehow a desired new value $T_+ > \bar{t}$ of the penalty parameter.
2. Linearize the left-hand-side $G_{t_+}(X, S)$ of the system of nonlinear equations at the point (\bar{X}, \bar{S}) , and replace (5) with the linearized system of equations

$$G_{t_+}(\bar{X}, \bar{S}) + \frac{\partial G_{t_+}(\bar{X}, \bar{S})}{\partial X}(X - \bar{X}) + \frac{\partial G_{t_+}(\bar{X}, \bar{S})}{\partial S}(S - \bar{S}) = 0. \quad (6)$$

3. Compute the corrections $\Delta X, \Delta S$ from the requirement that the updated pair $X_+ = \bar{X} + \Delta X$, $S_+ = \bar{S} + \Delta S$ must satisfy (4) and (6). In other words, the corrections should solve the system

$$\begin{aligned} \Delta X &\in \mathcal{L}, \\ \Delta S &\in \mathcal{L}^\perp, \\ G_{t_+}(\bar{X}, \bar{S}) + \frac{\partial G_{t_+}(\bar{X}, \bar{S})}{\partial X} \Delta X + \frac{\partial G_{t_+}(\bar{X}, \bar{S})}{\partial S} \Delta S &= 0 \end{aligned}$$

4. Complete updating the current solution by setting

$$\begin{aligned} X_+ &= \bar{X} + \Delta X, \\ S_+ &= \bar{S} + \Delta S. \end{aligned}$$

The primal-dual interior point methods we are describing basically fit the outlined scheme. However, if the current iterate (\bar{X}, \bar{S}) is not close enough to $Z_*(\bar{t})$, or the desired improvement $t_+ - \bar{t}$ is too large, the corrections given by the outlined scheme may be too large. As a result, the updating may be inappropriate. The standard way to overcome this difficulty is to replace the updating step with

$$X_+ = \bar{X} + \alpha \Delta X, \quad (7)$$

$$S_+ = \bar{S} + \beta \Delta S, \quad (8)$$

and to choose the stepsizes $\alpha > 0, \beta > 0$ from additional safety considerations, like ensuring the updated pair (X_+, S_+) to reside in the interior of \mathbf{K} , or enforcing it to stay in a desired neighborhood of the central path, etc.

5.2 Speed of path-tracing

In the LP-CQP-SDP situation, the speed at which the best, from the theoretical viewpoint, path-following methods manage to trace the path is inversely proportional to the square root of the parameter $\theta(K)$ of the underlying canonical barrier. This means that when started at a point (t^0, X^0, S^0) from the neighborhood $\mathcal{N}_{0.1}$ of the central path, the method after $O(1)\sqrt{\theta(K)}$ steps reaches the point $(t^1 = 2t^0, X^1, S^1) \in \mathcal{N}_{0.1}$, and so on.

Thus, it takes a fixed number $O(1)\sqrt{\theta(K)}$ steps to increase the current value of the penalty parameter by factor 2, staying all the time in $\mathcal{N}_{0.1}$. It then follows that every $O(1)\sqrt{\theta(K)}$ steps of the method reduce the (upper bound on the) inaccuracy of current approximate solutions by factor 2 or, which is the same, add a fixed number of accuracy digits to these solutions. Thus, the cost of an accuracy digit for the (best) path-following methods is $O(1)\sqrt{\theta(K)}$ steps.

5.3 Primal and dual path-following methods

The complexity analysis of the primal path-following method can be summarized in the following theorem:

Theorem 5.1. *Let $0 < \chi \leq \kappa \leq 0.1$. Assume that we are given a starting point (t_0, x_0, S_0) such that $t_0 > 0$ and the point*

$$(X_0 = \mathcal{A}x_0 - B, S_0)$$

is κ -close to $Z_(t_0)$:*

$$\text{dist}((X_0, S_0), Z_*(t_0)) \leq \kappa.$$

Starting with (t_0, x_0, X_0, S_0) , let us build the iterates (t_i, x_i, X_i, S_i) according to

$$\begin{aligned} t_i &= \left(1 + \frac{\chi}{\sqrt{\theta(K)}}\right) t_{i-1}, \\ x_i &= x_{i-1} + \underbrace{(-[\mathcal{A}^*(\nabla^2 K(X_{i-1}))\mathcal{A}]^{-1}[t_i c + \mathcal{A}^* \nabla K(X_{i-1})])}_{\Delta x_i}, \\ X_i &= \mathcal{A}x_i - B, \\ S_i &= -t_i^{-1}[\nabla K(X_{i-1}) + [\nabla^2 K(X_{i-1})\mathcal{A}\Delta x_i]]. \end{aligned}$$

The resulting process is well defined and generates strictly primal-dual feasible pairs (X_i, S_i) such that (t_i, X_i, S_i) stay in the neighborhood \mathcal{N}_k of the primal-dual central path.

The theorem says that, with properly chosen κ, χ (e.g., $\kappa = \chi = 0.1$), after we somehow reached the \mathcal{N}_k -neighborhood of the primal-dual central path, we can trace it by the primal path-following method, keeping the iterates in \mathcal{N}_k and increasing the penalty parameter by an absolute constant factor in every $O(\sqrt{\theta(K)})$ steps.

6 Complexity bounds for linear programming, conic quadratic programming, and semidefinite programming

6.1 Complexity of linear programming

Problem instance:

$$\min_{x \in \mathbb{R}^n} \{c^T x : a_i^T x \leq b_i, i = 1, \dots, m; \|x\|_\infty \leq R\}; \tag{p}$$

Data:

$$\begin{aligned}\text{Data}(p) &= [n; m; c; a_1, b_1; \dots; a_m, b_m; R], \\ \text{Size}(p) &\equiv \dim \text{Data}(p) = (m+1)(n+1) + 2.\end{aligned}$$

ϵ -solution: an $x \in \mathbb{R}^n$ such that

$$\begin{aligned}\|x\|_\infty &\leq R, \\ a_i^T &\leq b_i + \epsilon, i = 1, \dots, m, \\ c^T x &\leq \text{Opt}(p) + \epsilon\end{aligned}$$

Newton complexity of ϵ -solution:

$$\text{Compl}^{\text{Nwt}}(p\epsilon) = O(1)\sqrt{m+n}\text{Digits}(p, \epsilon),$$

where

$$\text{Digits}(p, \epsilon) = \ln \left(\frac{\text{Size}(p) + \|\text{Data}(p)\|_1 + \epsilon^2}{\epsilon} \right)$$

is the number of accuracy digits in ϵ -solution.

Arithmetic complexity of ϵ -solution:

$$\text{Compl}(p, \epsilon) = O(1)(m+n)^{3/2}n^2\text{Digits}(p, \epsilon).$$

6.2 Complexity of conic quadratic programming

Problem instance:

$$\min_{x \in \mathbb{R}^n} \{c^T x : \|A_i x + b_i\|_2 \leq c_i^T x + d_i, i = 1, \dots, m; \|x\|_2 \leq R\}; \quad (p)$$

Data:

$$\begin{aligned}\text{Data}(p) &= [n; m; k_1, \dots, k_m; c; A_1, b_1, c_1, d_1; \dots; A_m, b_m, c_m, d_m; R], \\ \text{Size}(p) &\equiv \dim \text{Data}(p) = (m + \sum_{i=1}^m k_i)(n+1) + m + n + 3.\end{aligned}$$

ϵ -solution: an $x \in \mathbb{R}^n$ such that

$$\begin{aligned}\|x\|_2 &\leq R, \\ \|A_i x + b_i\|_2 &\leq c_i^T x + d_i + \epsilon, i = 1, \dots, m, \\ c^T x &\leq \text{Opt}(p) + \epsilon\end{aligned}$$

Newton complexity of ϵ -solution:

$$\text{Compl}^{\text{Nwt}}(p\epsilon) = O(1)\sqrt{m+1}\text{Digits}(p, \epsilon).$$

Arithmetic complexity of ϵ -solution:

$$\text{Compl}(p, \epsilon) = O(1)(m+n)^{1/2}n \left(n^2 + m + \sum_{i=0}^m k_i^2 \right) \text{Digits}(p, \epsilon).$$

6.3 Semidefinite programming

Problem instance:

$$\min_{x \in \mathbb{R}^n} \left\{ c^T x : A_0 + \sum_{j=1}^n x_j A_j \succeq 0; \|x\|_2 \leq R \right\}, \quad (p)$$

where $A_j, j = 0, 1, \dots, n$, are symmetric block-diagonal matrices with m diagonal blocks $A_j^{(i)}$ of sizes $k_i \times k_i, i = 1, \dots, m$.

Data:

$$\begin{aligned} \text{Data}(p) &= [n; m; k_1, \dots, k_m; c; A_0^{(1)}, \dots, A_0^{(m)}; A_n^{(1)}, \dots, A_n^{(m)}; R], \\ \text{Size}(p) \equiv \dim \text{Data}(p) &= \left(\sum_{i=1}^m \frac{k_i(k_i + 1)}{2} \right) (n + 1) + m + n + 3. \end{aligned}$$

ϵ -solution: an $x \in \mathbb{R}^n$ such that

$$\begin{aligned} \|x\|_2 &\leq R, \\ A_0 + \sum_{j=1}^n x_j A_j &\succeq -\epsilon I, \\ c^T x &\leq \text{Opt}(p) + \epsilon \end{aligned}$$

Newton complexity of ϵ -solution:

$$\text{Compl}^{\text{Nwt}}(p\epsilon) = O(1) \left(1 + \sum_{i=1}^m k_i \right)^{1/2} \text{Digits}(p, \epsilon).$$

Arithmetic complexity of ϵ -solution:

$$\text{Compl}(p, \epsilon) = O(1) \left(1 + \sum_{i=1}^m k_i \right)^{1/2} n \left(n^2 + n \sum_{i=0}^m k_i^2 + \sum_{i=0}^m k_i^3 \right) \text{Digits}(p, \epsilon).$$

7 References

- [1] Aharon Ben-Tal and Arkadi Nemirovski. *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. SIAM, 2001. URL: <https://epubs.siam.org/doi/pdf/10.1137/1.9780898718829.fm>.
- [2] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004. URL: <https://stanford.edu/~boyd/cvxbook/>.
- [3] Visually Explained. *The Karush–Kuhn–Tucker (KKT) Conditions and the Interior Point Method for Convex Optimization*, Aug. 5, 2021 [Online]. URL: <https://www.youtube.com/watch?v=uh1Dk68cfWs&lc=UgxsvaU51ze2duBS3YR4AaABAg.9VSQAWQCXQX9VSXAj1PqJ6>.