Companion Report of PROXQP: an Efficient and Versatile Quadratic Programming Solver for Real-Time Robotics Applications and Beyond

Antoine Bambade, Fabian Schramm, Sarah El Kazdadi, Stéphane Caron, Adrien Taylor, Justin Carpentier

— Companion Report — ProxQP: an Efficient and Versatile Quadratic Programming Solver for Real-Time Robotics Applications and Beyond

Antoine Bambade[*,1,2], Fabian Schramm[1], Sarah El-Kazdadi[1], Stéphane Caron[1], Adrien Taylor[1] and Justin Carpentier[1]

A. Companion report organisation

This companion report provides the proof associated to the ProxQP algorithm introduced in [1]. It is organized as follows.

Section -B provides more details about which properties motivate the update rule presented in Section VI-B. Section -D establishes the global convergence of ProxQP (Theorem 1). It illustrates that BCL algorithm [2] finds appropriate proximal steps to match a desired convergence speed for primal feasibility p^k. More precisely, BCL decreases μ^k until the inequality $p^k \leqslant \epsilon_{\text{bcl}}^k$ holds for all iteration starting from k. At this stage of the algorithm, ProxQP amounts to a primal-dual proximal point algorithm, and we rely on [3, Proposition 1.2] to ensure the global convergence of ProxQP. We study the simplified Algorithm 2, which assumes that at each iteration k of the algorithm $\rho^k = \mu^k$ compared to Algorithm 1, where ρ is fixed. Updating ρ is not desirable in practice since it would require a numerical factorization at each iteration.

Section	Content
Section -B	More details about Section VI-B update rule
Section -C	Notations and technical properties
Section -D Missing technical elements.	Proof of Lemma 1
	Proof of Lemma 2
	Proof of Lemma 3
Section -E	Proof of Theorem 1
Section -F	Proof of Lemma 1

TABLE I: Organization of the companion report.

B. Details about the update rule presented in Section VI-B

The update rule presented in Section VI-B is motivated by the following intuition. Assume the steady state x_{goal} is still reachable by an unconstrained LQR scheme (see the grey line in Figure 1)

$$\begin{aligned}\min_{x_t\in\mathbb{R}^{n_x},u_t\in\mathbb{R}^{n_u}} & x_T^\top Q_T x_T + \sum_{t=0}^{T-1} x_t^\top Q x_t + u_t^\top R u_t, \\ & \text{s.t. } x_{t+1} = Ax_t + Bu_t,\ x_0 = x_{\text{init}},\end{aligned} \tag{1}$$

with a solution sequence $\{x_t^{\text{LQR}}\}_t$, $\{u_t^{\text{LQR}}\}_t$. As the constraints are linear, strong duality holds, and the primal of problem (1) p^{LQR} equals its dual δ^{LQR} for some dual multipliers $\{y_t^{\text{LQR}}\}_t$. Noting $\mathcal{L}^{\text{LQR}}$ the Lagrangian of problem (1), we have

$$\begin{aligned}-\infty < \delta^{\text{LQR}} &= \mathcal{L}^{\text{LQR}}(\{x_t^{\text{LQR}}\}_t, \{u_t^{\text{LQR}}\}_t, \{y_t^{\text{LQR}}\}_t) \\ &= \mathcal{L}(\{x_t^{\text{LQR}}\}_t, \{u_t^{\text{LQR}}\}_t, \{y_t^{\text{LQR}}\}_t, 0, 0), \\ &= \inf_{x_t,u_t} \mathcal{L}(\{x_t\}_t, \{u_t\}_t, \{y_t^{\text{LQR}}\}_t, 0, 0),\end{aligned}$$

with $\mathcal{L}$ the Lagrangian of the original problem (16), and 0 corresponds to the value of the dual multipliers associated with the box constraints with respect to x_t and u_t. It thus guarantees the dual of the original problem has a non-empty domain and that there exists a closest feasible solution sequence $\{\hat{x}_t\}_t, \{\hat{u}_t\}_t$ (see the orange line in Figure 1) with associated optimal shifts $\{\hat{s}_t^e\}_t, \{\hat{s}_t^x\}_t, \{\hat{s}_t^u\}_t$.

If we compare now $\|x_1^{\text{LQR}} - x_{\text{goal}}\|_2^2$ with $\|x_1 - x_{\text{goal}}\|_2^2$ assuming that both $x_1^{\text{LQR}} > \bar{x}$ and $\hat{x}_1 > \bar{x}$ we see

$$\begin{aligned}\|x_1^{\text{LQR}} - x_{\text{goal}}\|_2^2 = \|x_1^{\text{LQR}} - \bar{x}\|_2^2 + \|\bar{x} - x_{\text{goal}}\|_2^2 + \\ 2(x_1^{\text{LQR}} - \bar{x})^\top(\bar{x} - x_{\text{goal}}),\end{aligned}$$

and

$$\begin{aligned}\|x_1 - x_{\text{goal}}\|_2^2 &= \|x_1 - \bar{x}\|_2^2 + \|\bar{x} - x_{\text{goal}}\|_2^2 + 2(x_1 - \bar{x})^\top(\bar{x} - x_{\text{goal}}), \\ &= \|\hat{x}_1 - B\hat{s}_0^u - \bar{x}\|_2^2 + \|\bar{x} - x_{\text{goal}}\|_2^2 + \\ &\quad 2(\hat{x}_1 - B\hat{s}_0^u - \bar{x})^\top(\bar{x} - x_{\text{goal}}) \\ &= \|\underbrace{\hat{x}_1 - \bar{x}}_{=\hat{s}_1^x}\|_2^2 + \|B\hat{s}_0^u\|_2^2 - 2(B\hat{s}_0^u)^\top(\hat{s}_1^x) + \\ &\quad \|\bar{x} - x_{\text{goal}}\|_2^2 + 2(\hat{s}^x - B\hat{s}_0^u)^\top(\bar{x} - x_{\text{goal}}).\end{aligned}$$

Thus we have:

$$\begin{aligned}&\|x_1^{\text{LQR}} - x_{\text{goal}}\|_2^2 - \|x_1 - x_{\text{goal}}\|_2^2 = \|x_1^{\text{LQR}} - \bar{x}\|_2^2 + \\ &2(x_1^{\text{LQR}} - \bar{x})^\top(\bar{x} - x_{\text{goal}}) - \|\hat{s}_1^x\|_2^2 - \|B\hat{s}_0^u\|_2^2 - \\ &2(\hat{s}_1^x - B\hat{s}_0^u)^\top(\bar{x} - x_{\text{goal}}) + 2(B\hat{s}_0^u)^\top(\hat{s}_1^x).\end{aligned} \tag{2}$$

[*]Corresponding author. bambade.antoine@gmail.com
[1]Inria - Département d'Informatique de l'École normale supérieure, PSL Research University.
[2]École des Ponts, Marne-la-Vallée, France.

Since $x_1^{\text{LQR}} > \bar{x}$ and $\bar{x} > x_{\text{goal}}$, then $(x_1^{\text{LQR}} - \bar{x})^\top(\bar{x} - x_{\text{goal}}) > 0$. Hence, if the optimal shifts $\hat{s}_0^u$ and $\hat{s}_1^x$ are sufficiently small, Equation (2) shows that $\|x_1^{\text{LQR}} - x_{\text{goal}}\|_2 > \|x_1 - x_{\text{goal}}\|_2$. The same calculus shows in this setting that $\|x_1^{\text{LQR}} - \bar{x}\|_2 > \|x_1 - \bar{x}\|_2$. Consequently, this update rule guarantees in this setting that if x_{init} is not too far in the sense that the optimal shifts are small enough, then the update rule provides a point from which the feasible target x_{goal} is both closer to reach and closer of the feasible frontier. Our intuition is that the property of this update rule may eventually lead to a feasible update, with demonstration further investigated in future work.

C. Further notations and technical properties

This section provides a few additional technical ingredients that are necessary for establishing ProxQP convergence proof. The Fenchel convex conjugate of a function $\phi : \mathbb{R}^m \mapsto \mathbb{R}\cup\{\infty\}$ is defined as $\phi^\star : \mathbb{R}^m \mapsto \mathbb{R}\cup\{\infty\}$ though

$$\phi^\star(y) \stackrel{\text{def}}{=} \sup_{z\in\mathbb{R}^m} \{y^\top z - \phi(z)\} = -\inf_{z\in\mathbb{R}^m} \{\phi(x) - y^\top z\}. \quad (3)$$

We introduce a maximal monotone operator $T_{\mathcal{L}}^{-1}(v,t)$ (see for more details [4]), that encodes the set of solutions of the shifted problem (QP) by (t,v):

$$T_{\mathcal{L}}(x,z) \stackrel{\text{def}}{=} \{(v,t) | (v,-t) \in \partial\mathcal{L}(x,z)\},$$
$$T_{\mathcal{L}}^{-1}(v,t) \stackrel{\text{def}}{=} \arg\min_{x\in\mathbb{R}^n} \max_{z\in\mathbb{R}^m} \{\mathcal{L}(x,z) - x^\top v + z^\top t\},$$

where $\partial\mathcal{L}$ refers to

$$\partial\mathcal{L}(x,z) \stackrel{\text{def}}{=} \begin{pmatrix} \nabla f(x) + C^\top z \\ Cx - \partial I_u^\star(z) \end{pmatrix}, \quad (4)$$

where I_u denotes the indicator function of the constraint set $Cx \leqslant u$. The Lagrangian function and $T_{\mathcal{L}}(x,z)$ encodes the saddle sub-differential for (QP) [5, Section 4.3]. $T_{\mathcal{L}}$ and $T_{\mathcal{L}}^{-1}$ are generally "point to set" operators. From $T_{\mathcal{L}}$ and for $\mu > 0$ we also introduce the resolvent of $T_{\mathcal{L}}$ and $T_{\mathcal{L}}^{-1}$

- the resolvent of $T_{\mathcal{L}}$ [4]:
$$P_\mu \stackrel{\text{def}}{=} (I + \frac{1}{\mu} T_{\mathcal{L}})^{-1}, \quad (5)$$
- the resolvent of $T_{\mathcal{L}}^{-1}$ [4]:
$$Q_\mu \stackrel{\text{def}}{=} (I + \mu T_{\mathcal{L}}^{-1})^{-1} = I - P_\mu. \quad (6)$$

More precisely, when $T_{\mathcal{L}}^{-1}(0,0)$ is non-empty, it encodes the set of solutions to (QP). When (QP) is feasible, the KKT conditions for (QP) form a polyhedral variational inequality [6, Section 3D]. Hence, $\exists a \geqslant 0, \tau > 0$, such that $\forall (u,v) \in \mathbb{R}^{n+m}$:

$$\begin{gathered} \|(u,v)\| \leqslant \tau \\ \Longrightarrow \\ \text{dist}_{T_{\mathcal{L}}^{-1}(0,0)}(x,z) \leqslant a\|(u,v)\|, \forall (x,z) \in T_{\mathcal{L}}^{-1}(u,v) \end{gathered} \quad (7)$$

a and τ are constants describing the geometry of the QP. Rockafellar [7, Thm. 2] has established that it holds as soon as $T_{\mathcal{L}}(0,0)^{-1} \neq \emptyset$. That is, $\forall \mu > 0$, $\forall (x,z) \in \mathbb{R}^n \times \mathbb{R}^m$:

$$\begin{gathered} \mu\|Q_\mu(x,z)\| \leqslant \tau \\ \Longrightarrow \\ \text{dist}_{T_{\mathcal{L}}(0,0)^{-1}}(P_\mu(x,z)) \leqslant \frac{a\mu}{\sqrt{1+a^2\mu^2}} \text{dist}_{T_{\mathcal{L}}(0,0)^{-1}}(x,z). \end{gathered} \quad (8)$$

D. Convergence of ProxQP

The main convergence result for ProxQP is provided by Theorem 1 . For proving this theorem (in Section -E), we proceed with the following strategy. This strategy relies on a slightly different formulation of ProxQP provided by Algorithm 2 (it is necessary for the proof to consider the conceptual infinite sequence of iterates generated by the algorithm).

- First, subsubsection -D1 shows that the "if condition" of ProxQP (see Algorithm 2) is satisfied at infinitely many iterations.
- Second, subsubsection -D2 is more accurate and shows that the "else condition" of ProxQP (see Algorithm 2) is entered only for a finite number of iterations.
- It follows from the second point that ProxQP corresponds to an appropriate type of approximate primal-dual proximal point method, and hence inherits all its nice convergence properties. This is formally stated in Section -E.
- Finally, one can conveniently deduce from this construction that ProxQP asymptotically corresponds to a fixed-step proximal method of multiplier (PMM, i.e., there exists $K > 0$ such that for all $k \geqslant K$, μ^k is constant). This is formally stated in Section -F.

1) We enter the "if" of Algorithm 2 an infinite number of times

Lemma 1. *Under Assumption 1 and Assumption 2 , $\forall k \in \mathbb{N}, \exists K \geqslant 0$ such that the condition $\|[Cx^{k+K+1} - u]_+\|_\infty \leqslant \epsilon_{bcl}^{k+K}$ of Algorithm 2 is satisfied.*

Proof. The following proof proceeds by contradiction. That is, we show that there cannot exist a $k \geqslant 0$ such that $\forall K > 0$ it holds that $p^{k+K} \stackrel{\text{def}}{=} \|[Cx^{k+K+1} - u]_+\|_\infty > \epsilon_{\text{bcl}}^{k+K}$ (as defined in (8)). For convenience, and without loss of generality, we assume $k = 0$.

By contradiction hypothesis, we thus have that $p^K > \epsilon_{\text{bcl}}^K$ holds $\forall K \geqslant 0$. Therefore, $\mu^K = \mu^0(\mu_f)^K$ and $\epsilon_{\text{bcl}}^K = \epsilon_{\text{bcl}}^0(\mu^0)^{\alpha_{bcl}}(\mu_f)^{\alpha_{\text{bcl}}K}$ and hence $\mu^K \to 0$ and $\epsilon_{\text{bcl}}^K \to 0$ as $K \to \infty$, since $\mu_f \in (0,1)$ and $\alpha_{\text{bcl}} \in (0,1/2)$. For establishing the desired statement, we show that under the contradiction hypothesis, one has

$$\epsilon_{\text{bcl}}^K < \|[Cx^{K+1} - u]_+\|_\infty \leqslant A\sqrt{\mu^K}, \quad (9)$$

for all $K \geqslant 0$, thereby reaching a contradiction. Indeed, with $\alpha_{\text{bcl}} \in (0,1/2)$, the term $\sqrt{\mu^K}$ goes faster to zero (as a function of K) than ϵ_{bcl}^K. More precisely, the hyperparameter choices within ProxQP ensure that $\exists K_0 \geqslant 0$ such that $\forall K \geqslant K_0$, $\epsilon_{\text{bcl}}^K > A\sqrt{\mu^K}$.

For proving our claim, we, therefore, focus on proving (9) in the following lines.

First, at iteration $K+1$, the pair $(x^{K+1}, \hat{z}^{K+1})$ is an approximate solution to (10) . In other words, by defining intermediary variables, we have:

$$u^K \stackrel{\text{def}}{=} \nabla f(x^{K+1}) + C^\top \hat{z}^{K+1} + \mu^K(x^{K+1} - x^K),$$
$$v^K \stackrel{\text{def}}{=} [Cx^{K+1} - u + \mu^K(z^0 + (\alpha - 1)\hat{z}^{K+1}]_+ - \hat{z}^{K+1}\alpha\mu^K,$$

with $\|(u^K, v^K)\|_\infty \leqslant \epsilon^K$. Following [8, Section 3.1], one can equivalently look at $(x^{k+1}, \hat{z}^{k+1})$ as the unique minimum of the strongly convex function $\hat{\mathcal{M}}^K_{\rho,\mu^K,\alpha}$:

$$\begin{aligned}\hat{\mathcal{M}}^K_{\rho,\mu^K,\alpha}(x,z) \stackrel{\text{def}}{=} f(x) + \tfrac{\mu^K}{2}\|x-(x^K+\tfrac{1}{\mu^K}u^K)\|_2^2 \\ + \tfrac{1}{2\alpha\mu^K}\|[Cx-u+\mu^K(z^K+(\alpha-1)z)]_+\|_2^2 \\ + \tfrac{(1-\alpha)\mu^K}{2}\|z+\tfrac{1}{\alpha\mu^k}v^K\|_2^2.\end{aligned}$$

Therefore, for any $x \in \mathbb{R}^n$ satisfying $Cx - u \leqslant 0$ (such a feasible x exists by Assumption 2), we get:

$$\hat{\mathcal{M}}^K_{\rho,\mu^K,\alpha}(x^{K+1}, \hat{z}^{K+1}) \leqslant \hat{\mathcal{M}}^K_{\rho,\mu^K,\alpha}(x, 0).$$

As $\{x^K\}$ is bounded by Assumption 2 , then $\exists M > 0$ independent of K such that

$$\hat{\mathcal{M}}^K_{\rho,\mu^K,\alpha}(x^{K+1}, \hat{z}^{K+1}) \leqslant f(x) + M\mu^K, \tag{10}$$

where we used the facts that for a feasible point x, $[Cx - u + z^0\mu^K]_+ \leqslant [z^0]_+\mu^K$ (as $Cx - u \leqslant 0$), and that $\|(u^K, v^K)\|_\infty \leqslant \epsilon^K$ implies that $\|u^K\|_\infty/\mu^K \leqslant \epsilon^0$ and that $\|v^K\|_\infty/\mu^K$.

It follows from the definition of $\hat{\mathcal{M}}^K_{\rho,\mu^K,\alpha}$ (and nonnegativity of the $\|.\|_2$ terms) that:

- we can bound
$$\begin{aligned}\hat{\mathcal{M}}^K_{\rho,\mu^K,\alpha}(x^{K+1}, \hat{z}^{K+1}) \geqslant f(x^{K+1}) \\ + \tfrac{1}{2\alpha\mu^k}\|[Cx^{K+1} - u + \mu^k(z^0 + (\alpha-1)\hat{z}^{k+1})]_+\|_2^2\end{aligned}$$
which, using (10), can be used to reach
$$\begin{aligned}\|[Cx^{K+1} - u + \mu^K(z^0 + (\alpha-1)\hat{z}^{K+1})]_+\|_2 \\ \leqslant \sqrt{2\mu^K\alpha(f(x) - f(x^{K+1}) + M\mu^K)}.\end{aligned} \tag{11}$$
From (11), since $\{x^K\}_K$ is bounded and f is a quadratic, we deduce that there exists some $M_1 > 0$ independent of K such that
$$\|[Cx^{K+1} - u + \mu^K(z^0 + (\alpha-1)\hat{z}^{K+1})]_+\|_2 \leqslant \sqrt{\mu^K}M_1. \tag{12}$$
- We can bound
$$\begin{aligned}\hat{\mathcal{M}}^K_{\rho,\mu^K,\alpha}(x^{K+1}, \hat{z}^{K+1}) \geqslant f(x^{K+1}) \\ + \tfrac{(1-\alpha)\mu^K}{2}\|z + \tfrac{1}{\alpha\mu^k}v^K\|_2^2.\end{aligned}$$
which, using (10), can be used to reach
$$\begin{aligned}\sqrt{\mu^K}\|\hat{z}^{K+1}\|_2 \leqslant \tfrac{\sqrt{2}}{\sqrt{1-\alpha}}(\sqrt{f(x) - f(x^{K+1}) + M\mu^K} \\ + \tfrac{\sqrt{1-\alpha}}{\sqrt{\alpha}}\|v^K\|_2).\end{aligned}$$
It shows that $\sqrt{\mu^K}\|\hat{z}^{K+1}\|_2$ is bounded since $\{x^K\}$ is bounded and f is a quadratic function.

We are now properly armed to guarantee that $\forall K > 0$:

$$\|[Cx^{K+1} - u]_+\|_\infty \leqslant A\sqrt{\mu^K}. \tag{13}$$

for some $A > 0$. Indeed, it holds component-wise

$$\begin{aligned}[Cx^{K+1} - u]_+ - \mu^K|(z^0 + (\alpha-1)\hat{z}^{K+1})| \\ \leqslant [Cx^{K+1} - u + \mu^K(z^0 + (\alpha-1)\hat{z}^{K+1})]_+,\end{aligned} \tag{14}$$

where $|.|$ is the absolute value taken component-wise. We can deduce from (14) that

$$\begin{aligned}&\|[Cx^{K+1} - u]_+\|_2 \\ &\leqslant \|[Cx^{K+1} - u + \mu^K(z^0 + (\alpha-1)\hat{z}^{K+1})]_+\|_2 \\ &+ \sqrt{\mu^K}\|\sqrt{\mu^K}(z^0 + (\alpha-1)\hat{z}^{K+1})\|_2\end{aligned} \tag{15}$$

Hence, from (15), using (12) and the fact that $\{\sqrt{\mu^K}(z^0+(\alpha-1)\hat{z}^{K+1})\}_K$ is bounded by some $M_2 > 0$ (since $\{\sqrt{\mu^K}\hat{z}^{K+1}\}$ is bounded and $\{\sqrt{\mu^K}\}_K$ converges to zero) we get

$$\|[Cx^{K+1} - u]_+\|_2 \leqslant (\underbrace{M_1 + M_2}_{\stackrel{\text{def}}{=} A > 0})\sqrt{\mu^K}.$$

To reach the desired claim, we use (13) together with the contradiction hypothesis that $\epsilon^K_{\text{bcl}} < \|[Cx^{K+1} - u]_+\|_\infty$ which allows reaching the contradiction that $\forall K > 0$:

$$\epsilon^K_{\text{bcl}} \leqslant A\sqrt{\mu^K}.$$

□

2) We enter the “else” of Algorithm 2 a finite number of times

As previewed in Section -D, this section establishes that the “else” condition of Algorithm 2 is only entered a finite number of times. The proof can be summarized as follows.

- If the “else” condition is never satisfied, then the result directly holds.
- Otherwise, since Lemma 1 ensures the “if” condition is entered an infinite number of times, we consider the following setting. Without loss of generality, we can assume that the “else” condition is entered at iteration k for the Kth time and that it is followed by a “if” condition. It implies that
$$\mu^k = \mu^0(\mu_f)^K,\ \epsilon^k_{\text{bcl}} = \epsilon^0_{\text{bcl}}(\mu^k)^{\alpha_{\text{bcl}}}.$$
We denote by $N \geqslant 2$ the number of consecutive iterations from k for which we enter the “if” condition, meaning that
$$\mu^{k+N} = \mu^0(\mu_f)^K,\ \epsilon^{k+N-1}_{\text{bcl}} = \epsilon^0_{\text{bcl}}(\mu^k)^{\alpha_{\text{bcl}}+(N-1)\beta_{\text{bcl}}}.$$
In this situation, one can establish that for QPs, there exists constants $M > 0$ and $b \geqslant 1$ (describing the geometry of the problem at hand) such that the primal feasibility can be upper-bounded for $N \geqslant 2$ by
$$\|[Cx^{k+N} - u]_+\|_\infty \leqslant M\left(b\mu^k\right)^{N-1}, \tag{16}$$
For entering the “else” condition in this scenario at the $(k+N)$th step, it is therefore necessary to be in the situation where
$$\epsilon^{k+N-1}_{\text{bcl}} < \|[Cx^{k+N} - u]_+\|_\infty \leqslant M\left(b\mu^k\right)^{N-1},$$
Substituting the expressions for $\epsilon^{k+N-1}_{\text{bcl}}$ and μ^k, we arrive to the following necessary condition for entering the “else” at iteration $(k+N)$:
$$\epsilon^0_{\text{bcl}}(\mu^0(\mu_f)^K)^{\alpha_{\text{bcl}}+(N-1)\beta_{\text{bcl}}} \leqslant M(b\mu^0(\mu_f)^K)^{N-1}.$$

On the contrary, it is therefore sufficient to satisfy

$$M(b\mu^0(\mu_f)^K)^{N-1} \leqslant \epsilon^0_{\text{bcl}}(\mu^0(\mu_f)^K)^{\alpha_{\text{bcl}}+(N-1)\beta_{\text{bcl}}}$$

for not entering the "else". Given that $\alpha_{\text{bcl}} \in (0, 1/2)$, $\beta_{\text{bcl}} \in (0, 1)$ with $\beta_{\text{bcl}} + \alpha_{\text{bcl}} < 1$, and $\mu_f \in (0, 1)$ it can be established that this condition is met for all $N \geqslant 2$ as soon as

$$\max\left(\left\lceil \frac{\log(\frac{b(\mu^0)^{1-\alpha_{\text{bcl}}-\beta_{\text{bcl}}}M}{\epsilon^0_{\text{bcl}}})}{-\log(\mu_f)(1-\alpha_{\text{bcl}}-\beta_{\text{bcl}})} \right\rceil, \left\lceil \frac{\log(b(\mu^0)^{1-\beta_{\text{bcl}}})}{-\log(\mu_f)(1-\beta_{\text{bcl}})} \right\rceil, K_\tau\right) \leqslant K, \tag{17}$$

for some $K_\tau \in \mathbb{N}$ (for the pair (a, τ) from (8), which describes the geometry of the QP at hand). It shows that the "else" condition is entered only a finite number of times, thereby arriving to the desired conclusion.

Our main target is thus to establish that (17) holds, as the desired conclusion nicely follows from this fact. For reaching this target, Lemma 2 provides an intermediary result which ensures that for K sufficiently large (i.e., larger than K_τ mentioned in (17)), the technical property (8) holds for all the iterates of PROXQP.

Lemma 2. *Under Assumption 1, Assumption 2, and if $\mu^k \to 0$ (i.e., we enter an infinite number of times the "else" condition), for any $\tau > 0$ $\exists K_\tau \in \mathbb{N}$ such that for all $\forall k \geqslant K_\tau$, $\mu^k\|Q_{\mu^k}(x^k, z^k)\|_2 \leqslant \tau$.*

Proof. In this setting $\epsilon^k = \epsilon^0\mu^k$. Then, by definition of Q_{μ^k} (6)

$$\begin{aligned} \|Q_{\mu^k}(x^k, z^k)\|_2 &= \|(x^k, z^k) - P_{\mu^k}(x^k, z^k)\|_2 \\ &\leqslant \left\|\begin{pmatrix} x^k - x^{k+1} \\ z^k - z^{k+1}\end{pmatrix}\right\|_2 + \|(x^{k+1}, z^{k+1}) - P_{\mu^k}(x^k, z^k)\|_2. \end{aligned} \tag{18}$$

$P_{\mu^k}(x^k, z^k)$ corresponds to the solution pair to (3). Since $\mathcal{M}_{\mu^k,\mu^k,\alpha}$ is strongly convex with parameter at least $(1-\alpha)\mu^k$, it follows that

$$\begin{aligned} &(1-\alpha)\mu^k\|(x^{k+1}, z^{k+1}) - P_{\mu^k}(x^k, z^k)\|_2 \\ &\leqslant \|\nabla\mathcal{M}_{\mu^k,\mu^k,\alpha}(x^{k+1}, z^{k+1}) - \underbrace{\nabla\mathcal{M}_{\mu^k,\mu^k,\alpha}(P_{\mu^k}(x^k, z^k))}_{=0}\|_2 \\ &\leqslant \max(1, \tfrac{1-\alpha}{\alpha})\epsilon^k, \end{aligned}$$

where the last inequality comes from the fact that

$$\begin{aligned} &\nabla_z\mathcal{M}_{\mu^k,\mu^k,\alpha}(x^{k+1}, z^{k+1}) \\ &= \tfrac{1-\alpha}{\alpha}([Cx^{k+1} - u + \mu^k(z^k + (\alpha-1)z^{k+1})]_+ - \alpha\mu^k z^{k+1}), \end{aligned}$$

and (x^{k+1}, z^{k+1}) satisfies (11) at accuracy ϵ^k. Hence, it holds

$$\|(x^{k+1}, z^{k+1}) - P_{\mu^k}(x^k, z^k)\|_2 \leqslant \underbrace{\max(\tfrac{1}{1-\alpha}, \tfrac{1}{\alpha})}_{\stackrel{\text{def}}{=} w}\epsilon^k/\mu^k. \tag{19}$$

As $\epsilon^k = \epsilon^0\mu^k$, combining (18) with (19) allows obtaining the upper bound

$$\begin{aligned} \mu^k\|Q_{\mu^k}(x^k, z^k)\|_2 &\leqslant \mu^k\left\|\begin{pmatrix} x^k - x^{k+1} \\ z^k - z^{k+1}\end{pmatrix}\right\|_2 + w\epsilon^k \\ &\leqslant \mu^k\left\|\begin{pmatrix} x^k - x^{k+1} \\ z^k - z^{k+1}\end{pmatrix}\right\|_2 + \mu^k w\epsilon^0. \end{aligned} \tag{20}$$

Since $\{(x_k, z_k)\}$ is bounded (see Assumption 1), and since $\mu^k \to 0$ (exponentially in the number of times we enter the "else" condition), we conclude from (20) that $\forall\tau > 0$, there $\exists K_\tau \in \mathbb{N}$, such that $\forall k \geqslant K_\tau$, $\mu^k\|Q_{\mu^k}(x^k, z^k)\|_2 \leqslant \tau$. □

Finally, Lemma 3 establishes our main targets. More precisely, we make use of Lemma 2 to establish first the inequality (16) and then (17).

Lemma 3. *Under Assumption 1 and Assumption 2, $\exists N_{\max} \in \mathbb{N}$ such that for all $k \geqslant N_{\max}$ the condition $p^{k+1} \leqslant \epsilon^k_{bcl}$ of Algorithm 2 is satisfied (i.e., we enter the "if condition").*

Proof. We split the proof in several cases.

- If we enter the "else condition" a finite number of times, the result holds.
- If we enter the "else condition" an infinite number of times, Lemma 1 ensures that each time we enter the "else", there is a subsequent iteration for which we enter the "if".

  Therefore, we can without loss of generality assume that we enter the "else" at iteration k for the the Kth time, and that we enter the "if" at iteration $k + 1$.

  In this setup (we enter an infinite number of times the "else", $\mu^k \to 0$ and Lemma 2 applies. Therefore, we can also pick without loss of generality $k \geqslant K_\tau$ where $K_\tau \in \mathbb{N}$ is such that (8) holds (for a specific (a, τ) describing the geometry of the QP at hand).

  In what follows, we consider this situation and show that it cannot happen (i.e., "else" is entered a finite number of times).

Since (QP) is feasible by assumption, it follows that $T_{\mathcal{L}}^{-1}(0, 0) \neq \emptyset$. For $l \in [0, N]$ we denote by $(\bar{x}^{k+l}, \bar{z}^{k+l})$ the projection of (x^{k+l}, z^{k+l}) onto the set of solutions $T_{\mathcal{L}}^{-1}(0, 0)$. As $\bar{x}^{k+N}$ is feasible, it follows that

$$\|[Cx^{k+N} - u]_+\|_2 = \|[Cx^{k+N} - u]_+ - [C\bar{x}^{k+N} - u]_+\|_2. \tag{21}$$

As $[.]_+$ is a non-expansive operator (as any projection operator), it follows that (21) can be upper bounded by

$$\begin{aligned} \|[Cx^{k+N} - u]_+\|_2 &\leqslant \|C(x^{k+N} - \bar{x}^{k+N})\|_2 \\ &\leqslant \|C\|_2\|x^{k+N} - \bar{x}^{k+N}\|_2 \\ &\leqslant \|C\|_2 \underbrace{\left\|\begin{pmatrix} x^{k+N} - \bar{x}^{k+N} \\ z^{k+N} - \bar{z}^{k+N}\end{pmatrix}\right\|_2}_{\stackrel{\text{def}}{=}\text{dist}_{T_{\mathcal{L}}(0,0)^{-1}}(x^{k+N}, z^{k+N})}. \end{aligned} \tag{22}$$

The following lines therefore focus on upper bounding the last term in different ways. First,

$$\begin{aligned} &\left\|\begin{pmatrix} x^{k+N} - \bar{x}^{k+N} \\ z^{k+N} - \bar{z}^{k+N}\end{pmatrix}\right\|_2 \\ &\leqslant \left\|\begin{pmatrix} x^{k+N} \\ z^{k+N}\end{pmatrix} - \overline{P_{\mu^k}}(x^{k+N-1}, z^{k+N-1})\right\|_2, \end{aligned} \tag{23}$$

which follows from the definition of $(\bar{x}^{k+N}, \bar{z}^{k+N})$ as the projection of (x^{k+N}, z^{k+N}) and $\overline{P_{\mu^k}}(x^{k+N-1}, z^{k+N-1})$ corresponds to the projection of $P_{\mu^k}(x^{k+N-1}, z^{k+N-1})$ onto

$T_{\mathcal{L}}^{-1}(0,0)$. which can be further upper bounded via a triangle inequality:

$$\begin{aligned}&\| \begin{pmatrix} x^{k+N} \\ z^{k+N} \end{pmatrix} - \overline{P_{\mu^k}}(x^{k+N-1}, z^{k+N-1})\|_2 \\ &\leqslant \| \begin{pmatrix} x^{k+N} \\ z^{k+N} \end{pmatrix} - P_{\mu^k}(x^{k+N-1}, z^{k+N-1})\|_2 \\ &+ \|P_{\mu^k}(x^{k+N-1}, z^{k+N-1}) - \overline{P_{\mu^k}}(x^{k+N-1}, z^{k+N-1})\|_2.\end{aligned} \tag{24}$$

Hence we also have:

$$\begin{aligned}&\left\| \begin{pmatrix} x^{k+N} - \bar{x}^{k+N} \\ z^{k+N} - \bar{z}^{k+N} \end{pmatrix} \right\|_2 \\ &\leqslant \| \begin{pmatrix} x^{k+N} \\ z^{k+N} \end{pmatrix} - P_{\mu^k}(x^{k+N-1}, z^{k+N-1})\|_2 \\ &+ \|P_{\mu^k}(x^{k+N-1}, z^{k+N-1}) - \overline{P_{\mu^k}}(x^{k+N-1}, z^{k+N-1})\|_2.\end{aligned} \tag{25}$$

Now, using $k \geqslant K_\tau$ from Lemma 2 (for a specific (a, τ) such that (8) holds for the QP at hand), then using the property (8) provides

$$\begin{aligned}&\|P_{\mu^k}(x^{k+N-1}, z^{k+N-1}) - \overline{P_{\mu^k}}(x^{k+N}, z^{k+N})\|_2 \\ &\leqslant \frac{a\mu^k}{\sqrt{1 + a^2(\mu^k)^2}} \text{dist}_{T_{\mathcal{L}}(0,0)^{-1}}(x^{k+N-1}, z^{k+N-1}) \\ &\leqslant a\mu^k \left\| \begin{pmatrix} x^{k+N-1} - \bar{x}^{k+N-1} \\ z^{k+N-1} - \bar{z}^{k+N-1} \end{pmatrix} \right\|_2.\end{aligned} \tag{26}$$

Hence incorporating (26) in (25) leads to

$$\begin{aligned}&\left\| \begin{pmatrix} x^{k+N} - \bar{x}^{k+N} \\ z^{k+N} - \bar{z}^{k+N} \end{pmatrix} \right\|_2 \\ &\leqslant \tfrac{w\epsilon^{k+N-1}}{\mu^k} + a\mu^k \left\| \begin{pmatrix} x^{k+N-1} - \bar{x}^{k+N-1} \\ z^{k+N-1} - \bar{z}^{k+N-1} \end{pmatrix} \right\|_2\end{aligned} \tag{27}$$

where we also used the fact that (x^{k+N}, z^{k+N}) satisfies (11) at accuracy ϵ^{k+N-1} similarly to (19). Iteratively expanding the upper bound (27) $N-1$ times leads to

$$\begin{aligned}&\left\| \begin{pmatrix} x^{k+N} - \bar{x}^{k+N} \\ z^{k+N} - \bar{z}^{k+N} \end{pmatrix} \right\|_2 \\ &\leqslant w \sum_{l=0}^{N-1} \tfrac{\epsilon^{k+N-1-l}}{\mu^k} (a\mu^k)^l + (a\mu^k)^N \left\| \begin{pmatrix} x^{k} - \bar{x}^{k} \\ z^{k} - \bar{z}^{k} \end{pmatrix} \right\|_2.\end{aligned} \tag{28}$$

Because $\epsilon^{k+N-1-l} = \epsilon^0(\mu^k)^{N-l}$, we deduce from (28) that

$$\begin{aligned}&\left\| \begin{pmatrix} x^{k+N} - \bar{x}^{k+N} \\ z^{k+N} - \bar{z}^{k+N} \end{pmatrix} \right\|_2 \\ &\leqslant w(\mu^k)^{N-1} \sum_{l=0}^{N-1} (a)^l + (a\mu^k)^N \left\| \begin{pmatrix} x^{k} - \bar{x}^{k} \\ z^{k} - \bar{z}^{k} \end{pmatrix} \right\|_2 \\ &\leqslant w(\mu^k)^{N-1} \sum_{l=0}^{N-1} (a)^l + B(a\mu^k)^N,\end{aligned} \tag{29}$$

for some $B > 0$, since $\{(x^k, z^k)\}_k$ is bounded (by Assumption 2). Finally, incorporating (29) into (22) provides the bound

$$\|[Cx^{k+N} - u]_+\|_\infty \leqslant e\|C\|_2 \left(w(\mu^k)^{N-1} \sum_{l=0}^{N-1} (a)^l + B(a\mu^k)^N \right), \tag{30}$$

for some constant $e > 0$ (from the norm equivalences between $\|.\|_\infty$ and $\|.\|_2$ in $\mathbb{R}^m$). Depending on the value of a, we distinguish three different cases to show the inequality (16):

- if $a \in (0, 1)$, then, since $\{\mu^k\}_k$ is bounded by 1 (since $\mu^0 \in (0, 1)$ and $\mu_f \in (0, 1)$), we can upper bound the geometric sum in (30) with
$$\|[Cx^{k+N} - u]_+\|_\infty \leqslant \underbrace{e\|C\|_2 \left(w\tfrac{1}{1-a} + B \right)}_{=M} (\mu^k)^{N-1},$$
and hence (16) holds with $b = 1$.
- if $a > 1$, then we can upper bound the geometric sum with (using again that $\{\mu^k\}_k$ is bounded by 1)
$$\|[Cx^{k+N} - u]_+\|_\infty \leqslant \underbrace{e\|C\|_2 \left(w\tfrac{1}{a-1} + aB \right)}_{=M} (a\mu^k)^{N-1},$$
and hence (16) holds with $b = a$.
- if $a = 1$, then we can upper bound the geometric sum with (using again that $\{\mu^k\}_k$ is bounded by 1)
$$\begin{aligned}\|[Cx^{k+N} - u]_+\|_\infty &\leqslant e\|C\|_2 (wN + B) (\mu^k)^{N-1} \\ &\leqslant \underbrace{4e\|C\|_2 \max(w, B)}_{=M} (2\mu^k)^{N-1},\end{aligned}$$
where we used that $\forall N \in \mathbb{N}, N \leqslant 2^N$. Hence (16) holds with $b = 2$.

Thus in any case, there exists some $M > 0$ and $b \geqslant 1$, such that (16) holds. To conclude, in order to enter at least $N \geqslant 2$ consecutive times in the "if" condition it is sufficient to satisfy

$$M(b\mu^0(\mu_f)^K)^{N-1} \leqslant \epsilon^0_{\text{bcl}}(\mu^0(\mu_f)^K)^{\alpha_{\text{bcl}} + (N-1)\beta_{\text{bcl}}}, \tag{31}$$

which is equivalent to

$$\begin{aligned}&\log(\tfrac{b(\mu^0)^{1-\alpha_{\text{bcl}}-\beta_{\text{bcl}}} M}{\epsilon^0_{\text{bcl}}}) + (N-2)\log(b(\mu^0)^{1-\beta_{\text{bcl}}}) \\ &\leqslant -\log(\mu_f)(1 - \alpha_{\text{bcl}} - \beta_{\text{bcl}})K \\ &- \log(\mu_f)(1 - \beta_{\text{bcl}})(N-2)K.\end{aligned}$$

Since $\mu_f \in (0, 1)$, $\alpha_{\text{bcl}} \in (0, 1/2)$, $\beta_{\text{bcl}} \in (0, 1)$ with $\alpha_{\text{bcl}} + \beta_{\text{bcl}} < 1$, then (31) holds for any $N \geqslant 2$ as soon as

$$\max(\lceil \tfrac{\log(\frac{b(\mu^0)^{1-\alpha_{\text{bcl}}-\beta_{\text{bcl}}} M}{\epsilon^0_{\text{bcl}}})}{-\log(\mu_f)(1-\alpha_{\text{bcl}}-\beta_{\text{bcl}})} \rceil, \lceil \tfrac{\log(b(\mu^0)^{1-\beta_{\text{bcl}}})}{-\log(\mu_f)(1-\beta_{\text{bcl}})} \rceil, K_\tau) \leqslant K,$$

which establishes (17). In other words, we enter the "else condition" only a finite number of times (which is bounded above by the value of the previous max). □

E. Proof of Theorem 1

Theorem 1 (Convergence of PROXQP). *Under Assumption 1 (existence of a solution to (QP)), Assumption 2 (bounded iterates), and with all parameters set as required in the inputs of Algorithm 2, the iterates $\{(x^k, z^k)\}$ of* PROXQP *converge to a solution $(x^\star, z^\star)$ of (QP).*

Proof. The proof consists in showing that PROXQP asymptotically corresponds to a proximal point algorithm (PPA). Then, we leverage the classical global convergence guarantees for PPA, which is automatically inherited by PROXQP. To show this link between ProxQP and PPA, (i) we first recall PPA global convergence properties, and then (ii) show that in a finite number of iterations ProxQP is a fixed-step PPA.

When (QP) is feasible, the KKT conditions $0 \in T_{\mathcal{L}}(x,z)$ forms a polyhedral variational inequality [6, Section 3D]. In such settings, if we consider a sequence $\{\mu^k\}_k$, bounded below by some $\mu^\infty > 0$, and $\{\eta^k\}_k$ some summable sequence, then the inexact proximal point iteration

$$(x^{k+1}, z^{k+1}) \approx_{\eta^k} P_{\mu^k}(x^k, z^k), \tag{32}$$

with $\approx_{\eta^k}$ corresponding to

$$\|(x^{k+1}, z^{k+1}) - P_{\mu^k}(x^k, z^k)\|_2 \leqslant \eta^k, \tag{33}$$

is guaranteed to converge globally to some $(x^\star, z^\star) \in T_{\mathcal{L}}^{-1}(0,0)$ [3, Proposition 1.2].

Depending on the value of $k_{\max}$, we distinguish two different cases to show that ProxQP is asymptotically equivalent to PPA. As we will see, it amounts to show ProxQP always enter the "if" condition in a finite number of iterations.

- If $k_{\max} < +\infty$, then after $k_{\max}$ number of iterations, ProxQP iterates always enter the "if" condition.
- Otherwise, under Assumption 1 and Assumption 2 Lemma (3) ensures that there exists some $N_{\max} \in \mathbb{N}$, after which ProxQP enter only the "if" condition.

Hence, for $k \geqslant N \stackrel{\text{def}}{=} \min(k_{\max}, N_{\max})$, ProxQP always accepts the iterates $(x^{k+1}, \hat{z}^{k+1})$ from (10). Furthermore, for $l \geqslant 0$

$$\mu^{N+l} = \mu^N,$$
$$\epsilon^{N+l} = \epsilon^k (\mu^N)^l.$$

Hence the iterates of Algorithm 2 amount to a fixed step PPA, namely for $k \geqslant N$

$$(x^{k+1}, \hat{z}^{k+1}) \approx_{e\epsilon^k} P_{\mu^N}(x^k, z^k),$$

for some constant $e > 0$ which can be provided following the same strategy as detailed in (19). As $\{\epsilon^k\}_k$ is summable for $k \geqslant N$, it concludes our claim. □

F. Proof of Lemma 1

Lemma 4. *Under Assumption 1 (existence of solution to (QP)) and Assumption 2 (bounded iterates), and with all parameters set as required in the inputs of Algorithm 2,* $\exists K \in \mathbb{N}$, *and* $\mu \in \mathbb{R}$ *such that* $\forall k \geqslant K, \mu^k = \mu$. *Hence,* $\forall k \geqslant 0, p^{K+k} \leqslant \epsilon_{bcl}^K (\mu^K)^{\beta_{bcl} k}$.

Proof. By Lemma 3, $\exists \mu \in \mathbb{R}$, $\exists K \in \mathbb{N}$ such that for all $k \geqslant K$, it holds that $p^k < \epsilon_{\text{bcl}}^k$ and $\mu^k = \mu$ is constant. □

References

[1] A. Bambade, F. Schramm, S. El-Kazdadi, S. Caron, A. Taylor, and J. Carpentier, "PROXQP: an Efficient and Versatile Quadratic Programming Solver for Real-Time Robotics Applications and Beyond," 2023. [Online]. Available: https://inria.hal.science/hal-04133055

[2] A. R. Conn, N. I. M. Gould, and P. Toint, "A globally convergent augmented lagrangian algorithm for optimization with general constraints and simple bounds," *SIAM Journal on Numerical Analysis*, vol. 28, no. 2, pp. 545–572, 1991.

[3] F. J. Luque, "Asymptotic convergence analysis of the proximal point algorithm," *SIAM Journal on Control and Optimization*, vol. 22, no. 2, pp. 277–293, 1984.

[4] R. T. Rockafellar, "Monotone operators and the proximal point algorithm," *J. Control Optim.*, p. 877–898, 1976.

[5] E. K. Ryu and S. Boyd, "Primer on monotone operator methods," *Appl. comput. math*, vol. 15, no. 1, pp. 3–43, 2016.

[6] A. L. Dontchev, R. T. Rockafellar, and R. T. Rockafellar, *Implicit functions and solution mappings: A view from variational analysis*. Springer, 2009, vol. 616.

[7] R. T. Rockafellar, *Convex analysis*, ser. Princeton Mathematical Series. Princeton, N. J.: Princeton University Press, 1970.

[8] D. Marchi., "On a primal-dual newton proximal method for convex quadratic programs," *Computational Optimization and Applications*, vol. 76, pp. 451—-467, 2021.