A Proof of asymptotic convergence

This short proof consists in showing that the above assumptions satisfy conditions C1 through C5 of [21], as well as convergence of the non-stationary target to a fixed point. Our proof borrows elements from [24].

1. Since we have a convex program, the solution must be unique globally (see, for instance, Theorem 18 of [6]).
2. By convexity, second-order conditions are satisfied automatically. Constraint qualifications ensure that a linear approximation of the constraints properly characterizes the set of feasible search directions (and that strong duality holds [2]), thereby guaranteeing that the set of feasible descent directions is empty at $x^\star$. Strict complementarity avoids complications in stating sufficient optimality conditions. One sufficient condition for first-order constraint qualification is that the gradient vectors $\nabla c_i(x)$ restricted to active inequality constraints i are linearly independent. Under the above regularity assumptions, $\theta^\star$ is isolated and unique by second-order sufficiency (Theorem 4 of [6]) and, furthermore, by the Kuhn-Tucker Sufficiency Theorem [6], the unperturbed KKT conditions (4) are satisfied for some $z^\star \geq 0$.
3. Assumption 2 with $c_k = 1$ immediately implies Condition C1. The choice of c_k makes no difference in the gradient-based case; Spall [20] describes the role of c_k in the simultaneous perturbation algorithm. Note that the condition is applied to the maximum step sizes $\hat{a}_k$ (see Fig. 1), not the feasible step sizes $a_k \leq \hat{a}_k$. C2 is satisfied by the martingale difference assumption (see [20]).
4. Convexity of the objective $f(x)$ and constraints $c(x)$ implies convexity of the barrier function (3) because $-\log(-u)$ is convex and monotonically increasing in u [2]. By similar logic, the gradient of the barrier function is Lipschitz-continuous.
5. Theorem 25 of [6] or Theorem 3.12 of [7], convexity, strict complementarity, constraint qualifications and existence of the KKT solution $\theta^\star$ imply that unconstrained minimizers exist for all nonlinear systems $F_\mu(x, z) = 0$, the minimizer is isolated at $\mu = 0$, and every limit point of the time-varying target and barrier trajectory $\{x^\star_\mu\}$ converges in norm to $x^\star$ as long as $\sigma_k \to 0$. Due to continuity of the barrier function, condition C4 holds because the individual coordinates $x_{k,i}$ clearly satisfy $|x_{k+1,i} - x_{k,i}| < |x^\star_i - x_{k,i}|$ when k is large and when the iterate is bounded away from the solution (see [24])—and similarly for the components of z_k. C5 is guaranteed by having bounded iterates in combination with the other assumptions [20].
6. Convexity of the barrier function and boundedness of the gradient imply that the Newton step $(\Delta x_k, \Delta z_k)$ with gradient estimate y_k is sufficiently steep to push iterates toward $(x^\star_\mu, z^\star_\mu)$ (Condition C3).[4] However, this doesn't quite satisfy C3 as the step at iteration k is actually $(a_k \Delta x_k, a_k \Delta z_k)/\hat{a}_k$, so we need to show $\hat{a}_k > 0$ implies $a_k > 0$. Under the same assumptions of strict complementarity and first-order constraint qualification, El-Bakry *et al.* establish local convergence by deriving a bound on the step size, presented here with slight modification to fit the formulation of our algorithm:
$$a_k = \min\{1, \tau_k + O(\|F_\mu(x_k, z_k)\|) + O(\mu_k / \min_i(z_i c(x_i)))\}, \tag{14}$$
where $y = O(x)$ means that the sequence x bounds the sequence y from above [13], and τ_k is the largest number guaranteeing that a_k remains smaller than $\hat{a}_k$.[5] This identity applies without modification when $\nabla f(x_k)$ is replaced by y_k. Therefore, unbiased gradient estimates y_k will eventually produce a search direction with nonzero step sizes a_k, and condition C3 holds. The proof is complete.

---

[4] A differentiable function $f(x)$ is convex if and only if for all x, $f(x^\star) \geq f(x) + \nabla f(x)^T (x^\star - x)$ as shown in [2], which implies $(x - x^\star)\nabla f(x) \geq 0$.

[5] See: A. S. El-Bakry, R. A. Tapia, T. Tsuchiya, and Y. Zhang, *On the formulation and theory of the Newton interior-point method for nonlinear programming*, Journal of Optimization Theory and Applications, 89 (1996), pp. 507–541.

B Stability of primal-dual iterates near the central path

In this section, we take a few moments to step through Wright's [26] line of reasoning that establishes some vital stability properties of primal-dual iterates that closely follow the central path. The key points of [26] are as follows:

1. When solving for x in $Ax = b$ using a backward stable algorithm (e.g. the Cholesky factorization), the computed solution can be characterized as the exact solution to a system with perturbed matrix A and vector b [23]. Provided the iterate (x, z) is within a small δ-neighbourhood of the central path but not too close to the boundary of the feasible region (*i.e.* $c(x)$ is asymptotically bounded from below by δ), and $\mu = O(\delta)$, then the perturbation in the computed value of $b = -\nabla f_\mu(x)$ is on the order of ε, and the perturbation in $A = W - J^T \Sigma J$ is roughly proportional to $\epsilon_{\mathrm{machine}}/\delta$. Considering the poor condition number of A, this is not a good result. The next four points shed a more favourable light.
2. The matrix A is almost entirely dominated by a matrix that lies in the space spanned by the gradient vectors $\nabla c_i(x)$ of the active constraints i, and likewise for b. The perturbations of A and b are also restricted to this space. This is fortunate because this space corresponds precisely to a *well-conditioned invariant subspace of* A.
3. The (relative) condition numbers of the matrix-vector products Ax and $x = A^{-1}b$ are bounded from above by $\kappa(A) \equiv \|A\| \|A^{-1}\|$, where $\|A\|$ is an induced matrix norm of A [23]. The problem is that A—the matrix arising in the reduced system (6)—is notoriously ill-conditioned as μ approaches 0. Fortunately, we are able to derive a tighter bound by exploiting special structure within the matrix.
4. By dividing the singular values of a matrix into a group of large singular values and a group of small singular values ($A = U \Sigma V$ where Σ has diagonal sub-blocks Σ_{large} and Σ_{small}), the relative condition number of x projected onto the invariant subspace of Σ_{large} is bounded by $\kappa(\Sigma_{\mathsf{large}})$. When the separation of the singular values within each group is much smaller than the separation of the full set, then the bound $\kappa(\Sigma_{\mathsf{large}})$ is much tighter than the bound $\kappa(A)$.
5. The matrix in (6) possesses precisely these properties. Collecting all the evidence, the solution Δx projected onto the range space of active constraints has perturbations on the order of $\delta\varepsilon + \epsilon_{\mathrm{machine}}$, a factor of δ better than the bound stated above. The error bounds for the remaining portion of Δx are worse, $O(\delta\varepsilon + \epsilon_{\mathrm{machine}}/\delta)$, but it is precisely this portion—the projection onto the the null space of the Jacobian of the active constraints—that has negligible impact on determination of the first-order KKT conditions near the solution [13, Theorem 12.1]. (Note these error bounds are of little use when J is ill-conditioned.)
6. Finally, Wright [26] applies analogous analysis to the dual search direction to show that the computed value of Δz in (7) has a similar condition number.

In conclusion, an off-the-shelf backward stable Cholesky factorization can be used to solve (6) so long as the iterates are feasible.

Of concern is the ill-effect of *cancellation* in the constraints near $c_i(x) = 0$ (see Example 12.3 of [23]), but we defer the issue since cancellation did not arise in our experiments. Note that even though the full system (5) is well-conditioned, the solution is not any more accurate because the right-hand side incurs cancellation errors. S. J. Wright presents an alternate derivation of stability that does not require the linear independence constraint qualification.[6]

[6]See: S. J. Wright, *Effects of finite-precision arithmetic on interior-point methods for nonlinear programming*, SIAM Journal on Optimization, 12 (2001), pp. 36–78.