MIT 6.7220/15.084 — Nonlinear Optimization (Spring '25) May 6-8th 2025

Lecture 20
Central path and interior-point methods

Instructor: Prof. Gabriele Farina (✉ gfarina@mit.edu)*

Having laid the foundations of self-concordant functions, we are ready to see one of the most important applications of these functions: interior-point methods.

Since we will be working extensively with self-concordant functions, we will make the blanket assumption that Ω is an open, convex, nonempty set.

L20.1 Path-following interior-point methods: chasing the central path

Consider a problem of the form

$$\min_x \ \langle c, x\rangle \quad \text{s.t.} \ \ x \in \overline{\Omega},$$

where $c \in \mathbb{R}^n$ and $\overline{\Omega}$ denotes the closure of the open, convex, and nonempty set $\Omega \subseteq \mathbb{R}^n$.

Unlike iterative methods that *project* onto the feasible set (such as for example the projected gradient descent and the mirror descent algorithm), interior-point methods work by constructing a sequence of feasible points in Ω, whose limit is the solution to the problem. To do so, interior-point methods consider a sequence of optimization problems with objective

$$\gamma\langle c, x\rangle + f(x),$$

where $\gamma \geq 0$ is a parameter and f is a strongly nondegenerate self-concordant function on Ω.

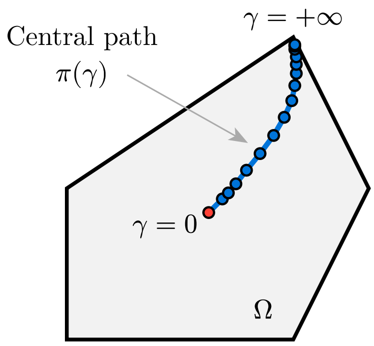

Figure: The central path traced by the sequence of solutions to the regularized problem $\arg\min\{-\gamma \cdot (x + y) + f(x) : x \in \Omega\}$, for increasing values of $\gamma \geq 0$. The self-concordant function f is the polyhedral barrier. The red dot, corresponding to the solution at $\gamma = 0$, is called *analytic center*.

As we saw in Lecture 19, self-concordant functions shoot to infinity at the boundary of their domain, and hence the minimizer of the self-concordant function will guarantee that the

*These notes are class material that has not undergone formal peer review. The TAs and I are grateful for any reports of typos.

solution is in the interior of the feasible set. The parameter γ is increased over time: as γ grows, the original objective function $\langle c, x\rangle$ becomes the dominant term, and the solution to the regularized problem will approach more and more the boundary. The path of solutions traced by the regularized problems is called the *central path*.

Definition L20.1 (Central path)**.** Let $f : \Omega \to \mathbb{R}$ be a lower-bounded strongly nondegenerate self-concordant function. The central path is the curve π parameterized over $\gamma \geq 0$, traced by the solutions[1] to the regularized optimization problem

$$\begin{aligned} \pi(\gamma) := \arg\min_{x} \quad & \gamma\langle c, x\rangle + f(x) \\ \text{s.t.} \quad & x \in \Omega. \end{aligned}$$

L20.1.1 Barriers and their complexity parameter

As it turns out, the performance of path-following interior-point methods depends crucially on a parameter of the strongly nondegenerate self-concordant function used, which is called the *complexity parameter* of the function.

Definition L20.2 (Complexity parameter)**.** The *complexity parameter* of a strongly nondegenerate self-concordant function $f : \Omega \to \mathbb{R}$ is defined as the supremum of the intrinsic squared norm of the second-order descent direction (Newton step) at any point in the domain, that is,

$$\theta_f := \sup_{x\in\Omega} \|n(x)\|_x^2.$$

Theorem L20.1 ([NN94], Corollary 2.3.3)**.** The complexity parameter of a strongly nondegenerate self-concordant function is at least 1.

We reserve the term *barrier* for only those self-concordant functions for which the complexity parameter is finite, as we make formal next.

Definition L20.3 (Barrier function)**.** A *strongly nondegenerate self-concordant barrier* (for us, simply *barrier*) is a strongly nondegenerate self-concordant function f whose complexity parameter is *finite*.

For example, in the case of the log barrier for the positive orthant, we can bound the complexity parameter as follows.

Example L20.1. The *logarithmic barrier* for the positive orthant $\mathbb{R}^n_{>0}$, defined as

[1] Remember that lower-bounded self-concordant functions always have a unique minimizer, as seen in Theorem L19.7 of Lecture 19.

$$f : \mathbb{R}^n_{>0} \to \mathbb{R} \qquad \text{where} \qquad f(x) = -\sum_{i=1}^n \log(x_i)$$

has complexity parameter $\theta_f = n$.

Solution. The Hessian of the logarithmic barrier is

$$\nabla^2 f(x) = \operatorname{diag}\left(\frac{1}{x_1^2}, ..., \frac{1}{x_n^2}\right),$$

and the Newton step is

$$n(x) = -[\nabla^2 f(x)]^{-1} \nabla f(x) = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Hence, the intrinsic norm of the Newton step satisfies

$$\|n(x)\|_x^2 = n(x)^\top [\nabla^2 f(x)] n(x) = \sum_{i=1}^n \frac{1}{x_i^2} x_i^2 = n$$

as we wanted to show. □

However, not all self-concordant functions are barriers.

Example L20.2. The function $f(x) = x - \log(x)$ is strongly nondegenerate self-concordant on $\Omega := \mathbb{R}_{>0}$, but it is not a barrier.

Solution. We already know that f is self-concordant, since $-\log(x)$ is self-concordant (see Lecture 19), and addition of linear functions to self-concordant functions preserve self-concordance.

The Hessian of f is $\nabla^2 f(x) = 1/x^2$, and the Newton step is correspondingly

$$n(x) = -[\nabla^2 f(x)]^{-1} \nabla f(x) = -x^2 \left(1 - \frac{1}{x}\right).$$

Hence, the intrinsic norm of the Newton step is

$$\|n(x)\|_x^2 = \frac{1}{x^2} \left[x^2 \left(1 - \frac{1}{x}\right)\right]^2 = x^2 - 2x + 1,$$

which is unbounded as $x \to +\infty$. □

L20.1.2 Complexity parameter and optimality gap of the central path

The complexity parameter of a barrier function is a crucial quantity that appears in the analysis of interior-point methods. We now begin with its first application in providing an upper bound on the optimality gap of the regularized problem.

Theorem L20.2. Let $f : \Omega \to \mathbb{R}$ be a barrier function. For any $\gamma > 0$, the point $\pi(\gamma)$ on the central path (see Definition L20.1), satisfies the inequality

$$\langle c, \pi(\gamma) \rangle \le \left(\min_{x' \in \overline{\Omega}} \langle c, x' \rangle \right) + \frac{1}{\gamma} \theta_f.$$

The above result ensures that when γ becomes large enough, then the points on the central path become arbitrarily close to the optimal value of the original problem. With little extra work, the same can be said for *approximate* solutions to $\pi(\gamma)$.

Theorem L20.3. Let $f : \Omega \to \mathbb{R}$ be a barrier function. For any $\gamma > 0$, and point $x \in \Omega$ such that $\|x - \pi(\gamma)\|_x \le \frac{1}{6}$,

$$\langle c, x \rangle \le \left(\min_{x' \in \overline{\Omega}} \langle c, x' \rangle \right) + \frac{6}{5\gamma} \cdot \theta_f.$$

L20.2 The (short-step) barrier method

The idea of the short-step barrier method is to chase the central path *closely* at every iteration. This is conceptually the simplest interior point method, with more advanced versions being the *long-step* barrier method and the *predictor-corrector* barrier method, which is what is implemented in commercial solvers such as CPLEX and Gurobi. We will use the term *short-step barrier method* and *barrier method* interchangeably today.

Assume that we know an *initial point* $x_1 \in \Omega$ that is close to the point $\pi(\gamma_1)$ on the central path, for some value of $\gamma_1 > 0$. The barrier algorithm now increases the parameter γ_1 to a value $\gamma_2 = \beta\gamma_1$ (where $\beta > 1$), and applies Newton's method to approximate the solution $\pi(\gamma_2)$. As long as x_1 was sufficiently close to $\pi(\gamma_1)$, we expect that in switching from γ_1 to γ_2, the point x_1 will still be in the region of quadratic convergence. In this case, Newton's method converges so fast, that (as we will see formally in the next subsection) a single Newton step is sufficient to produce a point $x_2 \coloneqq x_1 + n_{\gamma_2}(x_1)$ that is again very close to the central path at $\pi(\gamma_2)$. For the choice of parameter γ_2, the Newton step is in particular

$$x_2 \coloneqq x_1 - [\nabla^2 f(x_1)]^{-1} (\gamma_2 c + \nabla f(x_1)),$$

since the objective function we apply the second-order descent direction is by definition the problem

$$\begin{aligned} \min_x \quad & \gamma_2 \langle c, x \rangle + f(x) \\ \text{s.t.} \quad & x \in \Omega. \end{aligned}$$

Continuing this process indefinitely, that is,

$$\boxed{\gamma_{t+1} \coloneqq \beta\gamma_t, \qquad x_{t+1} \coloneqq x_t - [\nabla^2 f(x_t)]^{-1} (\gamma_{t+1} c + \nabla f(x_t))}$$

we have the *short-step barrier method.*

L20.2.1 Update of the parameter γ

As we did in Lecture 19, we will denote the second-order direction of descent—that is, the Newton step—starting from a point x using the letter n. However, since we are now dealing with a continuum of objective functions parameterized on γ, we will need to also specify what objective (that is, what value of γ) we are applying the Newton step to. For this reason, we will introduce the notation

$$n_\gamma(x) \coloneqq -[\nabla^2 f(x)]^{-1}(\gamma c + \nabla f(x)).$$

The main technical hurdle in analyzing the short-step barrier method is to quantify the proximity of the iterates to the central path. As is common with self-concordant functions, we will measure such proximity using the lengths of the Newton steps: *x_t is near $\pi(\gamma_t)$ in the sense that the intrinsic norm of the Newton step $n_{\gamma_t}(x_t)$ is small* (this should feel natural recalling Theorem L19.6 in Lecture 19).

How close to the central path is close enough, so that the barrier method using a single Newton update per iteration is guaranteed to work?

As we move our attention from the objective $\gamma_t\langle c, x\rangle + f(x)$ to the objective $\gamma_{t+1}\langle c, x\rangle + f(x)$, we can expect that distance to optimality of x_t to $\pi(\gamma_{t+1})$ increases by a certain amount compared to the distance from x_t to $\pi(\gamma_t)$. If this amount is not too large, then we can hope to use Theorem L19.8 in Lecture 19 to "recover" in a single Newton step the distance lost, and close the induction. The following theorem operationalizes the idea we just stated, and provides a concrete quantitative answer to what "close enough" means. In particular, we will show that $\|n_{\gamma_t}(x_t)\|_{x_t} \le \frac{1}{9}$ is enough.

Theorem L20.4. If x_t is close to the central path, in the sense that $\|n_{\gamma_t}(x_t)\|_{x_t} \le \frac{1}{9}$, then by setting

$$\gamma_{t+1} \coloneqq \beta\gamma_t \qquad \text{with} \quad \beta \coloneqq \left(1 + \frac{1}{8\sqrt{\theta_f}}\right),$$

the same proximity is guaranteed at time $t+1$, that is, $\|n_{\gamma_{t+1}}(x_{t+1})\|_{x_{t+1}} \le \frac{1}{9}$.

Proof. We need to go from a statement pertaining $\|n_{\gamma_t}(x_t)\|_{x_t}$ to one pertaining $\|n_{\gamma_{t+1}}(x_{t+1})\|_{x_{t+1}}$. We will do so in two steps.

■ **First part.** Observe the equality (valid for all γ_{t+1} and γ_t)

$$\begin{aligned} n_{\gamma_{t+1}}(x_t) &= -[\nabla^2 f(x_t)]^{-1}(\gamma_{t+1}c + \nabla f(x_t)) \\ &= -\frac{\gamma_{t+1}}{\gamma_t}[\nabla^2 f(x_t)]^{-1}\left(\gamma_t c + \frac{\gamma_t}{\gamma_{t+1}}\nabla f(x_t)\right) \\ &= -\frac{\gamma_{t+1}}{\gamma_t}[\nabla^2 f(x_t)]^{-1}(\gamma_t c + \nabla f(x_t)) + \frac{\gamma_{t+1}-\gamma_t}{\gamma_t}[\nabla^2 f(x_t)]^{-1}\nabla f(x_t) \\ &= \frac{\gamma_{t+1}}{\gamma_t}n_{\gamma_t}(x_t) + \left(\frac{\gamma_{t+1}}{\gamma_t} - 1\right)[\nabla^2 f(x_t)]^{-1}\nabla f(x_t). \end{aligned}$$

Using the triangle inequality for norm $\|\cdot\|_{x_t}$ and plugging in the hypotheses of the statement, we get

$$\begin{aligned}
\|n_{\gamma_{t+1}}(x_t)\|_{x_t} &\le \frac{\gamma_{t+1}}{\gamma_t}\|n_{\gamma_t}(x_t)\|_{x_t} + \left|\frac{\gamma_{t+1}}{\gamma_t} - 1\right| \cdot \|[\nabla^2 f(x_t)]^{-1}\nabla f(x_t)\|_{x_t} \\
&\le \frac{\gamma_{t+1}}{\gamma_t}\|n_{\gamma_t}(x_t)\|_{x_t} + \left|\frac{\gamma_{t+1}}{\gamma_t} - 1\right| \cdot \sqrt{\theta_f} \\
&\le \frac{1}{9}\left(1 + \frac{1}{8\sqrt{\theta_f}}\right) + \frac{1}{8\sqrt{\theta_f}}\sqrt{\theta_f} \\
&\le \frac{1}{9}\cdot\left(1 + \frac{1}{8}\right) + \frac{1}{8} = \frac{1}{4} \qquad \text{(since } \theta_f \ge 1\text{)}.
\end{aligned}$$

However, the left-hand side of the inequality is $\|n_{\gamma_{t+1}}(x_t)\|_{x_t}$ and *not* $\|n_{\gamma_{t+1}}(x_{t+1})\|_{x_{t+1}}$. This is where the second step comes in.

■ **Second part.** To complete the bound, we will convert from $\|n_{\gamma_{t+1}}(x_t)\|_{x_t}$ to $\|n_{\gamma_{t+1}}(x_{t+1})\|_{x_{t+1}}$. To do so, remember that x_{t+1} is obtained from x_t by taking a Newton step. Hence, using Theorem L19.8 of Lecture 19, we have

$$\|n_{\gamma_{t+1}}(x_{t+1})\|_{x_{t+1}} \le \left(\frac{\|n_{\gamma_{t+1}}(x_t)\|_{x_t}}{1 - \|n_{\gamma_{t+1}}(x_t)\|_{x_t}}\right)^2 \le \left(\frac{\frac{1}{4}}{1-\frac{1}{4}}\right)^2 = \frac{1}{9}.$$

This completes the proof. □

Remark L20.1. Remarkably, a safe increase in γ depends only on the complexity parameter θ_f of the barrier, and not on any property of the function. For example, for a linear program

$$\begin{aligned}
\min_x \quad & \langle c, x\rangle \\
\text{s.t.} \quad & Ax = b \\
& x \in \mathbb{R}^n_{\ge 0},
\end{aligned}$$

using the polyhedral barrier function, the increase in γ is independent of the number of constraints of the problem or the sparsity of A, and we can increase $\gamma_{t+1} = \gamma_t \cdot \left(1 + \frac{1}{8\sqrt{n}}\right)$.

The result in Theorem L20.4 shows that at every iteration, it is safe to increase γ by a factor of $1 + \frac{1}{8\sqrt{\theta_f}} > 1$, which leads to an exponential growth in the weight given to the objective function of the problem.

Hence, combining the previous result with Theorem L20.2 we find the following guarantee.

Theorem L20.5. Consider running the short-step barrier method with a barrier function f with complexity parameter θ_f, starting from a point x_1 close to $\pi(\gamma_1)$, *i.e.*, $\|n_{\gamma_1}(x_1)\|_{x_1} \le 1/9$, for some $\gamma_1 > 0$. For any $\varepsilon > 0$, after

$$T = \left\lceil 10\sqrt{\theta_f}\log\left(\frac{6\theta_f}{5\varepsilon\gamma_1}\right)\right\rceil$$

iterations, the solution computed by the short-step barrier method guarantees an ε-suboptimal objective value $\langle c, x_T\rangle \le (\min_{x\in\overline{\Omega}}\langle c, x\rangle) + \varepsilon$.

Proof. Since at every time the value of γ is increased by the quantity $1 + \frac{1}{8\sqrt{\theta_f}}$, the number of iterations required to increase the value from γ_1 to any value γ is given by

$$\begin{aligned} T &= \left\lceil \frac{\log\left(\frac{\gamma}{\gamma_1}\right)}{\log\left(1+\frac{1}{8\sqrt{\theta_f}}\right)} \right\rceil \\ &\le \left\lceil \log\left(\frac{\gamma}{\gamma_1}\right)\frac{5}{4}\cdot 8\sqrt{\theta_f}\right\rceil \qquad \left(\text{since } \frac{1}{\log(1+x)} \le \frac{5}{4x} \text{ for all } 0 \le x \le \frac{1}{2}\right) \\ &= \left\lceil 10\sqrt{\theta_f}\log\left(\frac{\gamma}{\gamma_1}\right)\right\rceil. \end{aligned}$$

On the other hand, we know from Theorem L20.3 that the optimality gap of $\pi(\gamma)$ is given by $6\theta_f/(5\gamma)$ as long as $\|x_T - \pi(\gamma_T)\|_{x_T} \le \frac{1}{6}$. This is indeed the case from Remark L19.3 of Lecture 19. So, to reach an optimality gap of ε, we need $\gamma = 6\theta_f/(5\varepsilon)$. Substituting this value into the previous bound yields the statement. □

L20.2.2 Finding a good initial point

The result in Theorem L20.5 shows that, as long as we know a point x_1 that is “close” (in the formal sense of Theorem L20.4) to the central path, for a parameter γ_1 that is not too small, then we can guarantee an ε-suboptimal solution in roughly $\sqrt{\theta_f}\log(1/\varepsilon)$ iterations.

■ **The analytic center.** Intuitively, one might guess that a good initial point for the algorithm would be a point close to $\zeta := \pi(0)$ (the minimizer of f on Ω), which is often called the *analytic center* of Ω. Let's verify that that is indeed the case. By definition, such a point satisfies $\nabla f(\zeta) = 0$, and so we have that

$$n_\gamma(\zeta) = -\gamma[\nabla^2 f(\zeta)]^{-1}c \qquad \Longrightarrow \qquad \|n_\gamma(\zeta)\|_\zeta = \gamma\cdot\|[\nabla^2 f(\zeta)]^{-1}c\|_\zeta.$$

Hence, $x_1 = \zeta$ is within proximity $1/9$ (in the sense of Theorem L20.4) of the central path for the value of

$$\gamma_1 = \frac{1}{9\,\|[\nabla^2 f(\zeta)]^{-1}c\|_\zeta}.$$

The only thing left to check is that γ_1 is not excessively small, so that the number of iterations predicted in Theorem L20.5 is not too large. We now show that indeed we can upper bound $\|[\nabla^2 f(\zeta)]^{-1}c\|_\zeta$.

Theorem L20.6. Let ζ be the minimizer of the barrier f on Ω. Then,

$$\|[\nabla^2 f(\zeta)]^{-1} c\|_\zeta \leq \langle c, \zeta\rangle - \min_{x \in \overline{\Omega}} \langle c, x\rangle.$$

(So, in particular, $\left\|[\nabla^2 f(\zeta)]^{-1} c\right\|_\zeta \leq \max_{x\in\overline{\Omega}} \langle c, x\rangle - \min_{x\in\overline{\Omega}} \langle c, x\rangle$.)

Proof. The direction $-[\nabla^2 f(\zeta)]^{-1} c$ is a descent direction for c, since

$$\left\langle c, -[\nabla^2 f(\zeta)]^{-1} c\right\rangle = -\left\|[\nabla^2 f(\zeta)]^{-1} c\right\|_\zeta^2 \leq 0.$$

Hence, as we consider points $x(\lambda) \coloneqq \zeta - \lambda \cdot [\nabla^2 f(\zeta)]^{-1} c$ for $\lambda \geq 0$ such that $x(\lambda) \in \Omega$, we have that the value of the objective $\langle c, x(\lambda)\rangle$ decreases monotonically, and in particular

$$\langle c, x(\lambda)\rangle = \langle c, \zeta\rangle - \lambda \cdot \left\|[\nabla^2 f(\zeta)]^{-1} c\right\|_\zeta^2,$$

which implies that

$$\left\|[\nabla^2 f(\zeta)]^{-1} c\right\|_\zeta^2 = \frac{\langle c, \zeta\rangle - \langle c, x(\lambda)\rangle}{\lambda} \leq \frac{\langle c, \zeta\rangle - \min_{x\in\overline{\Omega}}\langle c, x\rangle}{\lambda}.$$

To complete the proof, it suffices to show that we can move in the direction of $-[\nabla^2 f(\zeta)]^{-1} c$ for a meaningful amount λ. For this, we will use the property of self-concordant function that the Dikin ellipsoid $W(\zeta) \coloneqq \{x \in \Omega : \|x - \zeta\|_\zeta < 1\} \subseteq \Omega$. In particular, this implies that any $\lambda \geq 0$ such that

$$1 > \|\zeta - x(\lambda)\|_\zeta = \lambda \left\|[\nabla^2 f(\zeta)]^{-1} c\right\|_\zeta$$

generates a point $x(\lambda) \in \Omega$. So, we must have

$$\begin{aligned}\left\|[\nabla^2 f(\zeta)]^{-1} c\right\|_\zeta^2 &\leq \inf\left\{\frac{\langle c, \zeta\rangle - \min_{x\in\overline{\Omega}}\langle c, x\rangle}{\lambda} : 0 < \lambda < \frac{1}{\left\|[\nabla^2 f(\zeta)]^{-1} c\right\|_\zeta}\right\} \\ &= \left(\langle c, \zeta\rangle - \min_{x\in\overline{\Omega}}\langle c, x\rangle\right)\left\|[\nabla^2 f(\zeta)]^{-1} c\right\|_\zeta,\end{aligned}$$

which implies the statement. □

So, we have shown the following.

Theorem L20.7 (The analytic center ζ is a good initial point)**.** Let f be a barrier function with complexity parameter θ_f. If the short-step barrier method is initialized at the analytic center ζ, then the number of iterations required to obtain an ε-suboptimal solution is bounded by

$$T = \left\lceil 10\sqrt{\theta_f}\log\left(\frac{11\,\theta_f}{\varepsilon}\left(\langle c, \zeta\rangle - \min_{x\in\overline{\Omega}}\langle c, x\rangle\right)\right)\right\rceil.$$

■ Path switching and the auxiliary central path. In practice, we might not know where the analytic center is. In this case, the typical solution is to first approximate the analytic center, and then start the short step barrier method from there as usual.

To approximate the analytic center, one can use the *auxiliary central path*. The idea is the following: start from an arbitrary point $x' \in \Omega$. Such a point is on the central path traced by the solutions to

$$\pi'(\nu) := \arg\min_x -\nu\langle \nabla f(x'), x\rangle + f(x)$$
$$\text{s.t. } x \in \Omega.$$

Indeed, note that x' is the solution for $\nu = 1$, that is, $x' = \pi'(1)$.

We can then run the short-step barrier method chasing π' *in reverse*. At every step, we will *decrease* the value of ν by a factor of $1 - \frac{1}{8\sqrt{\theta_f}}$. Once the value of ν is sufficiently small that $\left\| [\nabla^2 f(x)]^{-1} \nabla f(x) \right\|_x \le 1/6$, we will have reached a point that is close to the analytic center, and we can start the regular short-step barrier method for $\pi(\gamma)$ from there. This technique is called *path switching*, since we follow two central paths (one from x' to the analytic center, and one from the analytic center to the solution), switching around the analytic center which path to follow. [▷ Try to work out the details and convince yourself this works!]

L20.3 Further readings

The short book by Renegar, J. [Ren01] and the monograph by Nesterov, Y. [Nes18] (Chapter 5) provide a comprehensive introduction to self-concordant functions and their applications in optimization.

I especially recommend the book by Renegar, J. [Ren01] for a concise yet rigorous account.

[Ren01] Renegar, J. (2001). *A Mathematical View of Interior-point Methods in Convex Optimization*. SIAM. https://doi.org/10.1137/1.9780898718812

[Nes18] Nesterov, Y. (2018). *Lectures on Convex Optimization*. Springer International Publishing. https://link.springer.com/book/10.1007/978-3-319-91578-4

Changelog

- Fixed a few typos (thanks Jonathan Huang!)
- Fixed a few typos (thanks Khizer!)