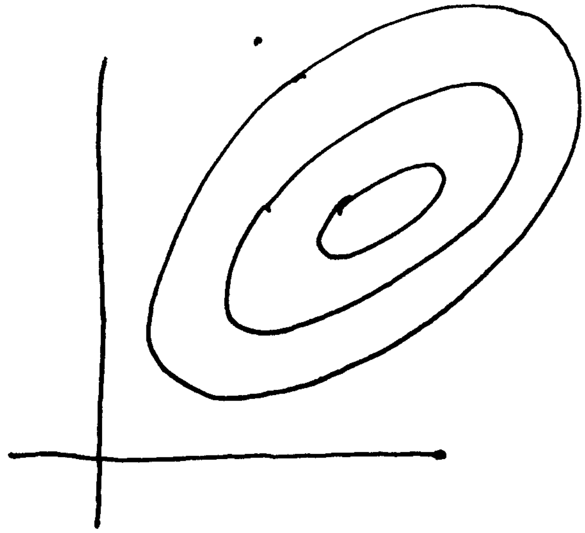
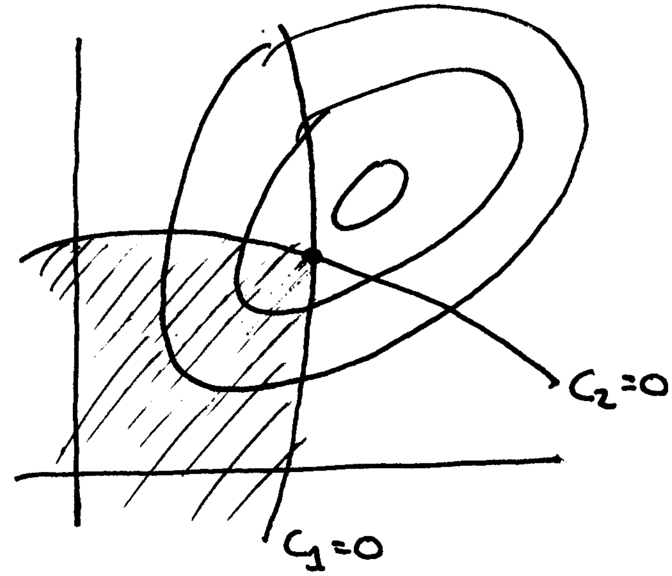
Constrained Minimization

We have now introduced some good ideas for minimizing $f: \mathbb{R}^n \to \mathbb{R}$.





In practice, it is very common to have a terrific solution that you can't use because it violates one or more constraints on the variables.

The general problems of this type are

$$\min f(\vec{x}) \quad \text{subject to } c_i(\vec{x})=0 \qquad\qquad \min f(\vec{x}) \quad \text{subject to } c_i(\vec{x})\geq 0$$

We call the region in $\mathbb{R}^n$ where the constraints are satisfied the <u>feasible</u> region.

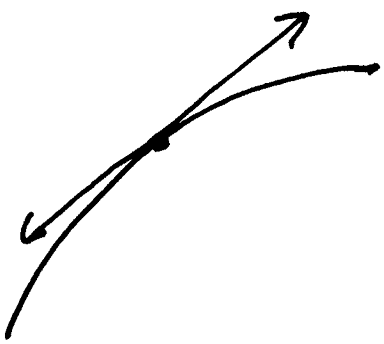
Our first goal is to derive conditions for a local minimizer x^* in the feasible region.

~~Suppose $\vec{x}^*+\vec{\delta}$ is feasible.~~ Then

$$\cancel{c_i(\vec{x}^*+\vec{\delta}) = c_i(\vec{x}^*) + \nabla c_i(\vec{x}^*)\cdot\vec{\delta} + o(|\vec{\delta}|)}.$$

~~where the $o(|\vec{\delta}|)$ terms are small with respect to $|\vec{\delta}|$ for small $\vec{\delta}$.~~

<u>Definition.</u> $\vec{s}$ is a feasible direction at $\vec{x}$ if $\vec{x}$ is feasible and $\exists$ a sequence of feasible points $\vec{x}_i \to \vec{x}$ so $\frac{\vec{x}_i - \vec{x}}{|\vec{x}_i - \vec{x}|} \to \vec{s}$.



equality constraints

inequality constraints

Proposition. If $\vec{x}^*$ is a local min for f in the feasible set, then there is no feasible direction $\vec{v}$ with $D_{\vec{v}} f < 0$.

Proof. Such a feasible direction would have some $\delta_i \to 0$ so that $\vec{\delta_i}/|\vec{\delta_i}| \to \vec{v}$ and $\vec{x}^* + \vec{\delta_i}$ feasible. By Taylor's Theorem, as $i \to \infty$,

$$f(\vec{x}^* + \vec{\delta_i}) = f(x^*) + \nabla f(x^*) \cdot \vec{\delta_i} + o(|\vec{\delta_i}|)$$

$$= f(x^*) + (D_{\vec{v}} f)\, |\vec{\delta_i}| + o(|\vec{\delta_i}|)$$

$$< f(x^*). \quad \times\!\!\!\times.$$

where we used implictly that $\frac{\delta_i}{|\delta_i|} - v \in o(|\delta_i|)$, for large enough i. $\square$

We now introduce a little more terminology.

Definition. A constraint c_i is <u>active</u> at x^* if $c_i(x^*) = 0$.

Note that equality constraints are

always active.

We let

$A =$ matrix whose columns are gradients of active constraints

(Weak) Kuhn-Tucker Theorem.

If A has full rank, then at x^*

$\nexists$ a feasible $\vec{v}$ with $D_{\vec{v}} f < 0$

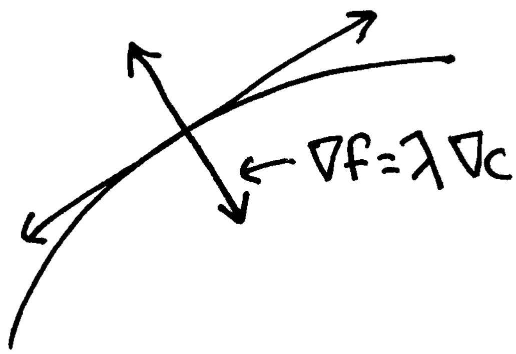
$$\Longleftrightarrow$$

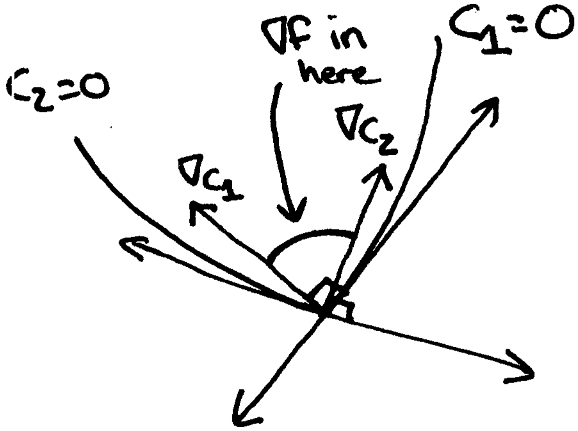
$\nabla f(x^*) = A\vec{\lambda}$, where $\lambda_i \geq 0$ if c_i is an inequality constraint

---

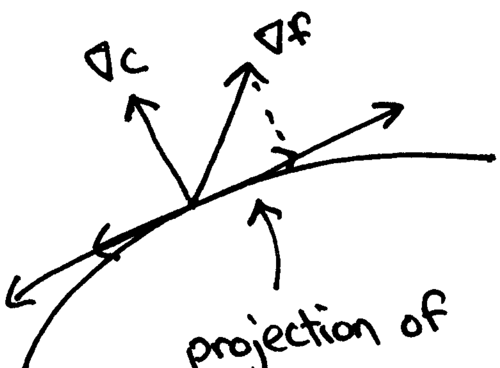
That is, there is no feasible direction which reduces f to first order $\Longleftrightarrow$ ∇f is a positive linear combination of ∇c_i

or ∇f is in the cone generated by the ∇c_i.

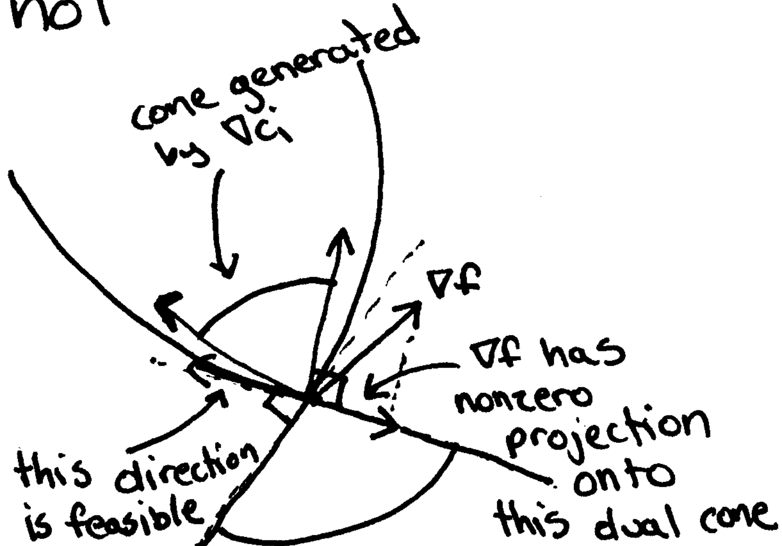




In each case, the proof is conceptually the same: suppose not

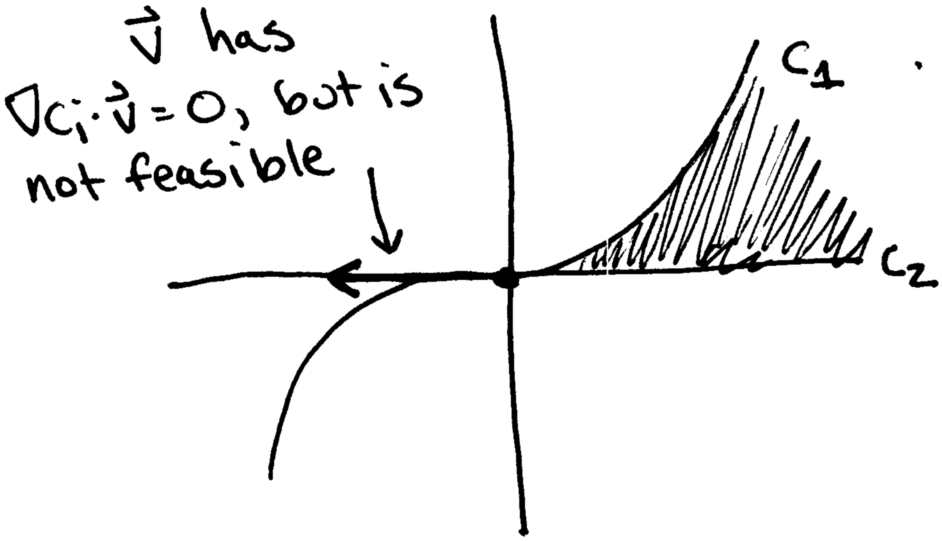


projection of ∇f to $\nabla c^{\perp} \neq 0$ $\Rightarrow$ f has a negative directional derivative in a feasible direction



↑ dual cone of vectors with $\nabla c_i \cdot v \leq 0$ for all i

We needed the assumption that A had full column rank to show that every vector v with $\nabla c_i \cdot v \geq 0$ for all i is actually a feasible direction:



this is called "constraint qualification".

Solving such a problem numerically is challenging, and there's no simple standard algorithm to pick. Here's one method, called "Rosen's projected gradient" method.

Projected Gradient Method.

1) Find active constraints, and construct the matrix A (generally sparse)

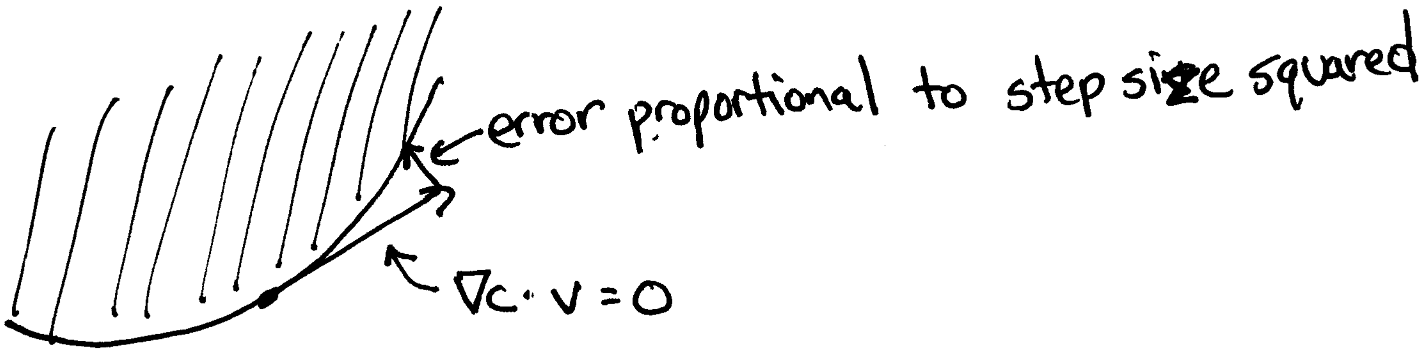
2) Find ∇f and the (positive) λ_i which minimize $|\nabla f - A\lambda|$. Construct the projected gradient $\nabla f - A\lambda = \nabla^P f$.

3) Line search in the $\nabla^P f$ diretion.

4) Correct error in constraints with Newton's method.

5) Repeat as needed.

Only step 4 needs some explanation. If the c_i are nonlinear functions, the fact that $\nabla^P f$ preserves the c_i to first order is not enough:



This has all the problems of steepest descent (and then some!) but it's pretty effective in practice.

Example. Ridgerunner. (25 mins)

Thank you!