

CONVEX OPTIMIZATION AND ENGINEERING APPLICATIONS (Formulary)

1. Introduction

Standard form for optimization problems $p^* = \min_x f_0(x)$ s.t. $f_i(x) \leq 0, i = 1, \dots, m$

where: $f_0 : \mathbb{R}^n \rightarrow \mathbb{R}$ (**objective function**), $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$, (inequality constraints)

Feasible set $\mathcal{X} = \{x \in \mathbb{R}^n : f_i(x) \leq 0\}$ — **Optimal solution** $x^* \in \mathcal{X} : f(x^*) = p^*$ — **Optimal value** p^*

Equality constraints $h_i(x) = 0, i = 1, \dots, p \iff h_i(x) \leq 0, -h_i(x) \leq 0$ — **Optimal Set** $\mathcal{X}_{opt} = \{x \in \mathcal{X} : f(x) = p^*\}$

Optimization problems in maximization form $p^* = \max_x f_0 \iff -p^* = \min_x -f_0(x) = \min_x g_0(x)$

ε -suboptimality of a solution $x \in \mathbb{R}^n \iff p^* \leq f_0(x) \leq p^* + \varepsilon$

2. Vector, Projections, Functions, Gradients

Vector $x \in \mathbb{R}^n = [x_1 \ x_2 \ \dots \ x_n]^T, x_i \in \mathbb{R}$ — **Sum** $x, y \in \mathbb{R}^n, x + y \iff x_i + y_i, i = 1, \dots, n$

Scalar multiplication $\alpha \in \mathbb{R}, x \in \mathbb{R}^n, \alpha x \iff \alpha x_i, i = 1, \dots, n$

Subspace $\mathcal{V} \subseteq \mathcal{X}$ is a **subspace** $\iff \forall x, y \in \mathcal{V}, \alpha x + \beta y \in \mathcal{V}$

$S = \{x^{(1)}, \dots, x^{(n)}\}, \text{span}(S) = \{\alpha_1 x^{(1)} + \dots + \alpha_n x^{(n)}, \alpha_i \in \mathbb{R}, i = 1, \dots, n\}$

Basis of a vector space $B = \{b^{(1)}, \dots, b^{(n)}\} \iff \text{span}(B) = \mathcal{X}, \forall x \in \mathcal{X}, x = \alpha_1 b^{(1)} + \dots + \alpha_n b^{(n)}, \alpha_i \in \mathbb{R}$

Direct sum of subspaces $\mathcal{X}, \mathcal{Y} \subseteq \mathbb{R}^n, \mathcal{X} \oplus \mathcal{Y} = \{x + y : x \in \mathcal{X}, y \in \mathcal{Y}\}$

Affine set \mathcal{X} vector space, $\mathcal{V} \subseteq \mathcal{X} \rightarrow$ affine set: $\mathcal{A} = \{x \in \mathcal{X} : x = x_0 + v, v \in \mathcal{V}\}$, note that: $\dim_{\mathbb{R}} \mathcal{A} = \dim_{\mathbb{R}} \mathcal{V}$

Line (1D-(affine set)) $L = \{x \in \mathbb{R}^n : x = x_0 + v, v \in \text{span}(u), \|u\|_2 = 1\}$

Norm $\|\cdot\| : \mathbb{R}^n \rightarrow \mathbb{R} \rightarrow \|x\| \geq 0, \|x\| = 0 \iff x = 0, \|xy\| = \|x\|\|y\|, \|\alpha x\| = |\alpha| \cdot \|x\|, \|x + y\| \leq \|x\| + \|y\|$

ℓ_p -norms $\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}, p = 2 \rightarrow$ Euclidean Distance, $p = 1 \rightarrow$ Manhattan Distance, $p = \infty \rightarrow \max x_i$

Inner Product $\langle x, y \rangle \geq 0, \langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle, \langle x, y \rangle = \langle y, x \rangle, \langle \alpha x, y \rangle = \alpha \langle x, y \rangle, \cos \theta = \frac{\langle x, y \rangle}{\|x\|_2 \|y\|_2}$

Standard inner product $x, y \in \mathcal{X}, \langle x, y \rangle = x^T y = \sum_{i=1}^n x_i y_i$, **Schwarz inequality** $\langle x, y \rangle \leq \|x\|_2 \|y\|_2$,

$\|x\|_2 = \sqrt{x^T x} = \sqrt{\langle x, x \rangle}, x \perp y \iff \langle x, y \rangle = 0 \iff \cos \theta = 0$

$S = \{x^{(1)}, \dots, x^{(n)}\}$ **mutually orthogonal** $\iff \langle x, y \rangle = \begin{cases} 0 & i \neq j \\ \neq 0 & i = j \end{cases}$ **orthonormal** $\iff \langle x, y \rangle = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$

Vector \perp **Subspace** $\mathcal{S} \subseteq \mathcal{X}, x \in \mathcal{X}, x \perp \mathcal{S} \iff \langle x, s \rangle = 0, \forall s \in \mathcal{S}, \mathcal{S}^\perp = \{x \in \mathcal{X} : \langle x, s \rangle = 0, \forall s \in \mathcal{S}\}$

Orthogonal decomposition of a vector $\forall x \in \mathcal{X}, x = x_1 + x_2 : x_1 \in \mathcal{S}, x_2 \in \mathcal{S}^\perp \iff \mathcal{X} = \mathcal{S} \oplus \mathcal{S}^\perp$

Projections $\Pi_{\mathcal{S}}(x) = \min_{y \in \mathcal{S}} \|y - x\|$ — (**Projection Theorem**) $\begin{cases} x^* \in \mathcal{S} \text{ is unique} \\ (x - x^*) \perp \mathcal{S} \iff (x - x^*) \in \mathcal{S}^\perp \end{cases}$

Functions $f : \mathbb{R}^n \rightarrow \mathbb{R}$ (function), $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ (map), **Domain of** f : $\text{dom}(f) = \{x \in \mathbb{R}^n : \|f(x)\| < \infty\}$

$\text{graph}(f) = \{(x, f(x)) \in \mathbb{R}^{n+1} : x \in \mathbb{R}^n\}, \text{epi}(f) = \{(x, t) \in \mathbb{R}^{n+1} : x \in \mathbb{R}^n, t \geq f(x)\}$

Contour curves $C_f(t) = \{x \in \mathbb{R}^n : f(x) = t, t \in \mathbb{R}\}$ — α -**sublevel set** $S_\alpha = \{x \in \mathbb{R}^n : f(x) \leq t, t \in \mathbb{R}\}$

f is **linear** $\iff f(\alpha x_1 + \beta x_2) = \alpha f(x_1) + \beta f(x_2), \alpha, \beta \in \mathbb{R}, \tilde{f}$ is **affine** $\iff \tilde{f} - x_0$ is linear

any affine function can be written as: $f(x) = a^T x + b$ where $a \in \mathbb{R}^n, b \in \mathbb{R}, b = f(0), a_i = f(e_i) - b$,

Hyperplane $\mathcal{H} = \{x \in \mathbb{R}^n : a^T x = b\}, a \in \mathbb{R}^n$ (normal direction), $\mathcal{H}_- = \{x \in \mathbb{R}^n : a^T x \leq b\}$, (open half-space)

$\mathcal{H}_{++} = \{x \in \mathbb{R}^n : a^T x > b\}$ (closed half-space)

Gradient Given $f : \mathbb{R}^n \rightarrow \mathbb{R}$ differentiable $\nabla f = \left[\frac{\partial f}{\partial x_1} \ \frac{\partial f}{\partial x_2} \ \dots \ \frac{\partial f}{\partial x_n} \right]^T, \nabla f(x_0)^T \cdot v$ (directional derivatives)

the rate of variation is maximal when $\nabla f(x_0) \parallel v$ minimal when $\nabla f(x_0) \perp v$, moreover $\nabla f(x_0) \perp C_f(t) \forall x_0 \in \text{dom}(f)$

3. Matrices

Matrix $A \in \mathbb{R}^{m,n}, a_{ij} \in \mathbb{R} — AB_{ij} = R_i(A) \cdot C_j(B)$ (rows by column) — $(AB)^T = B^T A^T$

An **affine map** $f = Ax + b, A \in \mathbb{R}^{m,n}, b \in \mathbb{R}^n$ (generalization of an **affine function**)

Subspaces associated with a matrix $\begin{cases} \mathcal{R}(A) = \{Ax : x \in \mathbb{R}^n\} & \text{range of } A \\ \mathcal{N}(A) = \{x \in \mathbb{R}^n : Ax = 0\} & \text{nullspace of } A \\ \dim_{\mathbb{R}} \mathcal{R}(A) = \text{rank}(A), 1 \leq \text{rank}(A) \leq \min(m, n) & \text{rank of } A \end{cases}$

Fundamental thm. of linear algebra For any $A \in \mathbb{R}^{m,n}$ $\begin{cases} \mathbb{R}^n \supseteq \mathcal{N}(A) \perp \mathcal{R}(A^T) \equiv \mathcal{N}(A)^\perp \iff \mathbb{R}^n = \mathcal{N}(A) \oplus \mathcal{R}(A^T) \\ \mathbb{R}^m \supseteq \mathcal{R}(A) \perp \mathcal{N}(A^T) \equiv \mathcal{R}(A)^\perp \iff \mathbb{R}^m = \mathcal{R}(A) \oplus \mathcal{N}(A^T) \\ \forall x \in \mathbb{R}^n, x = A^T x + \zeta, \zeta \in \mathcal{N}(A) \\ \forall w \in \mathbb{R}^m, w = Ax + \xi, \xi \in \mathcal{N}(A^T) \end{cases}$

Singular matrix $A \in \mathbb{R}^{n,n}$ is **singular** $\iff \det(A) = 0 \iff \mathcal{N}(A) \neq \{0\} \iff \text{rank}(A) \neq \min(m, n)$

Inverse matrix $A \in \mathbb{R}^{n,n}, \det(A) \neq 0 \exists A^{-1} \in \mathbb{R}^{n,n} : AA^{-1} = I_n, (AB)^{-1} = B^{-1} A^{-1}$

Similar matrices $A, B \in \mathbb{R}^{n,n}$ are similar $\iff \exists P : B = P^{-1}AP, P$ is **non-singular** (columns: basis for $\mathbb{R}^{n,n}$)

Eigenvalues/Eigenvector $u \in \mathbb{R}^{n,n}$ **eigenvector** for $A \iff Au = \lambda u, u \neq 0$,

$\lambda \rightarrow$ **eigenvalue** associated with u , **Eigenvalues** \rightarrow roots of $p(\lambda) = \det(A - \lambda I_n) = 0$, **Eigenspace** $\rightarrow \phi_i = \mathcal{N}(A - \lambda_i I_n)$

Algebraic Multiplicity ν_i (multiplicity of the root of $p(\lambda)$) **Geometric Multiplicity** $\mu_i = \dim_{\mathbb{R}}(\phi_i) \rightarrow \nu_i \leq \mu_i$

Diagonalizable matrices (thm.) $\lambda_i, i = 1, \dots, k \leq n$ (distinct eigenvalues of A), $U^{(i)} = [u_1^{(i)} \ \dots \ u_{\nu_i}^{(i)}]$ the eigenvector wrt λ_i , if $\nu_i = \mu_i, i = 1, \dots, k, U = [U^{(1)} \ \dots \ U^{(k)}]$ is **invertible** and $A = U \Lambda U^{-1}, \Lambda = \text{diag}(\lambda_1 I_{\mu_1}, \dots, \lambda_k I_{\mu_k})$

Matrix norms ...is a function $f : \mathbb{R}^{m,n} \rightarrow \mathbb{R}$ — additional property $f(AB) \leq f(A)f(B)$ **sub-multiplicativity**

Frobenius norm (extension of ℓ_2 -norm) $\|A\|_F = \sqrt{\text{trace}(A^T A)} = \sqrt{\sum_{j=1}^m \sum_{i=1}^n a_{ij}^2} = \sqrt{\sum_{i=1}^n \lambda_i(A^T A)}$

Operator norms (max input-output gain of $y = Au$) $\|A\|_p \doteq \max_{u \neq 0} \frac{\|Au\|_p}{\|u\|_p} = \max_{\|u\|_p=1} \|Au\|_p$ (ℓ_p -induced norms)

$\|A\|_1 = \max_{j=1,\dots,m} \sum_{i=1}^n |a_{ij}|$, $\|A\|_\infty = \max_{i=1,\dots,m} \sum_{j=1}^n |a_{ij}|$, $\|A\|_2 = \sqrt{\lambda_{\max}(A^T A)}$ (see variational characterization)

Spectral radius $A \in \mathbb{R}^{n,n}$ $\rho(A) \doteq \max_{i=1,\dots,n} |\lambda_i(A)|$ — $\rho(A) \leq \min(\|A\|_1, \|A\|_\infty)$

4. Symmetric Matrices and SVD

$A \in \mathbb{R}^{n,n}$ is **symmetric** if $AA^T = I_n \iff A^{-1} = A^T$, \mathbb{S}^n is the subspace of **symmetric matrices** $n \times n$

Examples: (Hessian matrix given $f : \mathbb{R}^2 \rightarrow \mathbb{R}$) $H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix}$, (quadratic function) $q(x) = \frac{1}{2}x^T Hx + c^T x + d$

Quadratic form Given $A \in \mathbb{S}^n$, $\forall x \neq 0$ the function $q(x) = x^T A x$ is a **quadratic form associated with A**

Spectral theorem Given $A \in \mathbb{S}^n$ it holds that: $\lambda_i \in \mathbb{R}$, $i = 1, \dots, n$, exists $U = [u_1 \dots u_n]$ orthogonal and $A = U \Lambda U^T = \sum_{i=1}^n \lambda_i u_i u_i^T$, $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$. I can sort them $\lambda_{\max}(A) = \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n = \lambda_{\min}(A)$

Rayleigh quotient given $A \in \mathbb{S}^n$, we can define $\forall x \in \mathbb{R}^n$ the quantity $r(x) = \frac{x^T A x}{x^T x}$ (Rayleigh quotient)

Theorem For any $A \in \mathbb{S}^n$, $\forall x \neq 0$, $\lambda_{\max} \leq r(x) \leq \lambda_{\min}$, where $\begin{cases} \lambda_{\max}(A) = \max_{\|x\|_2=1} x^T A x & (x^* = u_1) \\ \lambda_{\min}(A) = \min_{\|x\|_2=1} x^T A x & (x^* = u_n) \end{cases}$

...on the ℓ_2 -induced norm (by using the definition of operator norm and ℓ_2 -norm) $\frac{\|Ax\|_2^2}{\|x\|_2^2} = \frac{(Ax)^T (Ax)}{x^T x} = \frac{x^T (A^T A) x}{x^T x}$

that is the Rayleigh quotient of $A^T A \in \mathbb{S}^n$ from the theorem follows that $\|A\|_2 = \max_{\|x\| \neq 0} r_{A^T A}(x) = \sqrt{\lambda_{\max}(A^T A)}$ \square

Given a symmetric matrix, it is said to be $\begin{cases} \text{Positive definite}(A \succ 0) & \forall x \in \mathbb{R}^n, x^T A x > 0 \iff \lambda_i > 0, i = 1, \dots, n \\ \text{Positive semidefinite}(A \succeq 0) & \forall x \in \mathbb{R}^n, x^T A x \geq 0 \iff \lambda_i \geq 0, i = 1, \dots, n \\ \text{Negative definite}(A \prec 0) & \forall x \in \mathbb{R}^n, x^T A x < 0 \iff \lambda_i < 0, i = 1, \dots, n \\ \text{Negative semidefinite}(A \preceq 0) & \forall x \in \mathbb{R}^n, x^T A x \leq 0 \iff \lambda_i \leq 0, i = 1, \dots, n \end{cases}$

Matrix square-root $A \succeq 0 \Rightarrow \exists B : A = B^2$, $B = A^{1/2}$ is the **matrix square-root**

Cholesky decomposition $A \succ 0 \iff \exists B : A = B^T B$, such a B can be computed as: $B = U \Lambda^{1/2} U^T$, where U comes from the spectral decomposition and $\Lambda^{1/2} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n})$

SINGULAR VALUE DECOMPOSITION(SVD) (It holds for any matrix $A \in \mathbb{R}^{m,n}$, $A^T A$ is very important!)

Theorem Given $A \in \mathbb{R}^{m,n}$, it can be written as: $A = U \tilde{\Sigma} V^T$, $U \in \mathbb{R}^{m,m}$, $V \in \mathbb{R}^{n,n}$ are *orthogonal matrices* while

$\tilde{\Sigma} = \begin{bmatrix} \Sigma & 0_{r,n-r} \\ 0_{m-r,r} & 0_{m-r,n-r} \end{bmatrix}$, $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_r)$, $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$ are the **singular values**, $U = [u_1, \dots, u_r]$ are the **left singular vector**, $V = [v_1, \dots, v_r]$ are the **right singular vector**.

SVD (compact form) $A = U_r \Sigma V_r = \sum_{i=1}^r (\sigma_i u_i v_i^T)$, moreover $\begin{cases} \sigma_i^2 = \lambda_i(A^T A), \sigma_1^2 = \lambda_{\max}(A^T A) \\ u_i \text{ eigenvectors of } A A^T \\ v_i \text{ eigenvectors of } A^T A \end{cases}$

Properties of SVD We can redefine some properties by using SVD $\begin{cases} \text{rank}(A) = r \\ \mathcal{N}(A) = \mathcal{R}([v_{r+1} \dots v_n]), \mathcal{R}(A) = \mathcal{R}([u_1 \dots u_r]) \\ \|A\|_F^2 = \sum_{i=1}^r \sigma_i^2, \|A\|_2^2 = \lambda_{\max}(A^T A) = \sigma_1^2 \\ \|A\|_* = \sum_{i=1}^r \sigma_i \text{ nuclear norm} \end{cases}$

Moore-Penrose pseudo-inverse $A^\dagger = V \tilde{\Sigma}^\dagger U^T$ (or in compact form) $A^\dagger = V_r \Sigma^{-1} U_r^T$, (full column) $A^\dagger = (A^T A)^{-1} A$

Condition number How close is A to be singular? $\kappa(A) = \frac{\sigma_1}{\sigma_n}$ the larger κ the closest A to be singular.

Low rank approximation $A_k = \arg \min_{A_k \in \mathbb{R}^{m,n}, \text{rank}(A_k)=k} \|A - A_k\|_F^2 = \sum_{i=1}^k (\sigma_i u_i v_i^T)$

Ratio of the total variance in A_k is the quantity $\eta_k = \frac{\|A_k\|_F^2}{\|A\|_F^2} = \frac{\sigma_1^2 + \dots + \sigma_k^2}{\sigma_1^2 + \dots + \sigma_n^2}$ is useful a graph k vs η_k to choose

properly k — **Norm approximation error** $e_k = \frac{\|A - A_k\|_F^2}{\|A\|_F^2} = 1 - \eta_k$

Application: Principal component Analysis (PCA) I collect in a matrix $X \in \mathbb{R}^{n,m}$ by columns m samples

characterized by n features. $\tilde{X} = [x_1 - \bar{x} \dots x_m - \bar{x}]$, $\bar{x} = \frac{1}{m} \sum_{i=1}^m x_i$ (baricenter). $x_i \in \mathbb{R}^n$

I want a direction $z \in \mathbb{R}^n$ along which the projections of the centered data have the maximal variance.

Ingredients: $\begin{cases} \text{projection on } z & \alpha_i = \tilde{x}_i^T z, i = 1, \dots, m \\ \text{variance} & \frac{1}{m} \sum_{i=1}^m \alpha_i^2 = \sum_{i=1}^m z^T \tilde{x}_i \tilde{x}_i^T z = z^T \tilde{X} \tilde{X}^T z \end{cases}$

Problem to solve: $\max_{z \in \mathbb{R}^n} z^T \tilde{X} \tilde{X}^T z$, $\tilde{X} = U_r \Sigma V_r^T$, $\tilde{X} \tilde{X}^T = U_r \Sigma^2 U_r^T$ (spectral decomposition)

from the theorem: $z = u_1$ (**first principal component**) \rightarrow the i -th principal component $\iff i$ -th row of U from SVD. $U_k \in \mathbb{R}^{n,k}$ (first k principal components, columns of U) \rightarrow I choose k according to η_k , then: $x = \bar{x} + U_k z$, $z \in \mathbb{R}^k$ (zero-mean random factors). **Result:** we are using $z \in \mathbb{R}^k$, instead of $x \in \mathbb{R}^n$, $k < n$ to represent our data. \square

5. Least Squares (LS)

Applications: $\begin{cases} \text{approximation for a "fat" system of equations} \\ \text{polynomial approximation/regression} \end{cases} \quad y = Ax \xrightarrow{\text{relax}} y \simeq Ax, \text{ residuals } r(x) = Ax - y$

Objective: find an x such that the squared-residuals are "small" $f_0(x) = \sum_i r_i^2(x) = \sum_i (a^T x - y_i)^2 = \|Ax - y\|_2^2$

Least Squares problem: $\min_{x \in \mathbb{R}^n} \|Ax - y\|_2^2$ (unconstrained) — **Solution** we have to find the roots of $\nabla f_0(x) = 0$

Rewriting of the f_0 $\|Ax - y\|_2^2 = x^T Q x - 2b^T x + c$, $Q = A^T A \succ 0$, $b = A^T y$, $c = \|y\|_2^2$

$\nabla f_0(x) = 0 \iff Qx = b \iff \mathbf{A}^T \mathbf{A}x = \mathbf{A}^T \mathbf{y}$ (normal equations) $x = (A^T A)^{-1} A^T y$ (for full column A)

Geometric interpretation (projections) LS problem can be recasted as: $\min_{\tilde{y} \in \mathcal{R}(A)} \|\tilde{y} - y\|_2^2$

You know from the projection theorem that: (i) $\tilde{y} \in \mathcal{R}(A)$, (ii) $(\tilde{y} - y) \perp \mathcal{R}(A) \iff (\tilde{y} - y) \in \mathcal{N}(A^T) \iff A^T(\tilde{y} - y) = 0 \iff A^T(Ax - y) = 0 \iff A^T Ax = A^T y$ (normal equations)

Variants for the LS problem	LS+Equality constraints	$\min_x f_0$ s.t. $Cx = d \rightarrow x = \bar{x} + Nz$, $\text{span}(N) = \mathcal{N}(C)$
		$\min_x \ \bar{A}x - \bar{y}\ _2^2$, $\bar{A} = AN$, $\bar{y} = y - A\bar{x}$
	LS+weighted residuals	$f_0 = \sum w_i r_i^2(x) = \ W(Ax - y)\ _2^2$, $W = \text{diag}(w_1, \dots, w_n)$
		$\min_x \ A_w x - y_w\ _2^2$, $A_w = WA$, $y_w = Wy$
	LS+ ℓ_2 -regularization	$\min_x \ Ax - y\ _2^2 + \gamma \ x\ _2^2$, $\left\ \begin{bmatrix} a \\ b \end{bmatrix} \right\ _2^2 = \ a\ _2^2 + \ b\ _2^2$
		$\min_x \ \tilde{A}x - \tilde{y}\ $, $\tilde{y} = [y \ 0_n]^T$, $\tilde{A} = [A \ \sqrt{\lambda} I_n]^T$

6. Convexity: sets and functions

Given $P = \{x^{(1)}, x^{(2)}, \dots, x^{(m)}\}$, $\begin{cases} \text{linear combination} & \lambda_1 x^{(1)} + \lambda_2 x^{(2)} + \dots + \lambda_m x^{(m)}, \lambda_i \in \mathbb{R} \\ \text{convex combination} & \lambda_i \in \mathbb{R}, \lambda_i \geq 0, \sum \lambda_i = 1 \end{cases}$

Well-known subspaces $\begin{cases} \text{linear hull} & \text{subspace with all linear combinations} \\ \text{affine hull} & \text{aff}(P), \lambda_i \in \mathbb{R}, \sum \lambda_i = 1 \\ \text{convex hull} & \text{co}(P) \text{ subspace with all convex combinations} \\ \text{conic hull} & \text{conic}(P) \lambda_i \in \mathbb{R}, \lambda_i \geq 0 \end{cases}$

CONVEX SETS $C \subseteq \mathbb{R}^n$ **convex** $\iff \forall x, y \in C \rightarrow$ the segment joining x and y lies in C . More formally, a convex combination of any two points falls within C ...

$\forall x, y, \lambda \in [0, 1], x, y \in C \rightarrow \lambda x + (1 - \lambda)y \in C$, $x \neq y, \lambda \in (0, 1), \lambda x + (1 - \lambda)y \in \text{relint}(C)$

Cone and convex cone Given $C \subseteq \mathbb{R}^n$ is a **cone** $\iff x \in C \rightarrow \alpha x \in C, \alpha > 0$ — C also convex \rightarrow **convex cone**

Operations preserving convexity $\begin{cases} \text{Intersections with convex sets} & C_1, \dots, C_n \text{ convex} \rightarrow C = \bigcap_{i=1}^n C_i \text{ is convex} \\ \text{Affine transformations} & f(x) = Ax + b, C \text{ cvx} \rightarrow f(C) = \{f(x) : x \in C\} \text{ cvx} \end{cases}$

Supporting Hyperplane Given C convex set, $z \in \partial C$, \mathcal{H} **supporting hyperplane** $\iff z \in \mathcal{H}$ and $C \subseteq \mathcal{H}_-$
Thm.: exists separating \mathcal{H} for C at z

Separating Hyperplane Given C_1, C_2 convex sets the hyperplane \mathcal{H} : $\begin{cases} \text{separates the sets} & C_1 \subseteq \mathcal{H}_-, C_2 \subseteq \mathcal{H}_+ \\ \text{strictly separates the sets} & C_1 \subseteq \mathcal{H}_{--}, C_2 \subseteq \mathcal{H}_{++} \end{cases}$
Thm.: $C_1 \cap C_2 = \emptyset$ exists sep. \mathcal{H}

CONVEX FUNCTIONS

$f : \mathbb{R}^n \rightarrow \mathbb{R}$ is $\begin{cases} \text{convex} \iff \text{(i) } \text{dom}(f) \text{ convex, (ii) } x, y \in \text{dom}(f), \lambda \in [0, 1] \rightarrow f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) \\ \text{strictly convex} \iff x, y \in \text{dom}(f), \lambda \in [0, 1] \rightarrow f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y) \\ \text{strongly convex} \iff \exists m > 0 : \tilde{f}(x) = f(x) - \frac{m}{2} \|x\|_2^2 \text{ is convex} \xrightarrow{\text{imply}} \text{strictly convex} \end{cases}$

Properties f convex (cvx) $\iff \text{epi}(f)$ is cvx — f cvx $\Rightarrow S_\alpha$ (α -sublevel set) is convex $\forall \alpha$ — f str. cvx $\nRightarrow S_\alpha$ str. cvx

f, g strictly convex $\Rightarrow f + g$ strictly convex — f convex, g strongly convex $\Rightarrow f + g$ strongly convex

h, g convex, g non decreasing $\Rightarrow h \circ g$ cvx — g concave, h convex non-increasing $\Rightarrow h \circ g$ cvx

Non-negative linear combinations $f_i : \mathbb{R}^n \rightarrow \mathbb{R}, i = 1, \dots, m$ convex $\Rightarrow f = \sum_{i=1}^m \alpha_i f_i, \alpha_i \geq 0$ is convex over $\bigcap_i \text{dom}(f_i)$

Affine variable transformation $f : \mathbb{R}^n \rightarrow \mathbb{R}, g(x) = f(Ax + b)$ is convex

Conditions for convexity $\begin{cases} \text{First-order condition} & f \text{ convex} \iff \forall x, y \in \text{dom}(f), f(y) \geq f(x) + \nabla f(x)^T(y - x) \\ & f \text{ differentiable} \\ \text{Second-order condition} & \begin{cases} f \text{ convex} \iff \nabla^2 f(x) \succeq 0 \ \forall x \in \text{dom}(f) \\ f \text{ strictly convex} \iff \nabla^2 f(x) \succ 0 \\ f \text{ strongly convex} \iff \exists m : \nabla^2 f(x) \succ mI, \ \forall x \in \text{dom}(f) \end{cases} \\ & f \text{ twice differentiable} \\ \text{Restriction to a line} & f \text{ convex} \iff \underbrace{g(t) = f(x_0 + tv)}_{\text{restriction of } f \text{ to a line}}, \ x_0, v \in \mathbb{R}^n, \ t \in \mathbb{R} \text{ is convex} \end{cases}$

Pointwise maximum $f_\alpha(x)$ convex and $\alpha \in \mathcal{A}, \mathcal{A}$ compact set (closed and bounded), $f(x) = \max_{\alpha \in \mathcal{A}} f_\alpha(x)$ is convex

Jensen's inequality Given $f : \mathbb{R}^n \rightarrow \mathbb{R}$ convex and $z \in \mathbb{R}^n$ a random variable such that $p\{z \in \text{int dom}(f)\} = 1$ it holds that: $f(\mathbb{E}[z]) \leq \mathbb{E}[f(z)]$. **Case z discrete R.V.** $p(z = x^{(i)}) = \theta_i, i = 1, \dots, m, \sum_i \theta_i = 1, \theta_i \geq 0 \Rightarrow f(\sum_i \theta_i x^{(i)}) \leq \sum_i \theta_i f(x^{(i)})$

7. Convex problems

$$\begin{aligned} p^* &= \min_{x \in \mathbb{R}^n} \overbrace{f_0(x)}^{\text{convex}} \\ \text{s.t.} \quad & \underbrace{f_i(x) \leq 0}_{\text{convex} \rightarrow 0\text{-sublevel set}}, \ i = 1, \dots, m \\ & \underbrace{h_i(x) = 0, i = 1, \dots, p}_{\text{affine} \rightarrow \text{flat}} \end{aligned}$$

(1)-(3) \rightarrow **CONVEX OPTIMIZATION PROBLEM**

Feasible set \mathcal{X} : convex since intersection between convex sets. $\mathcal{X} = \mathbb{R}^n$ (**unconstrained problem**).

Active/inactive constraint $\begin{cases} f_i(x^*) < 0 & \text{inactive at } x^* \\ f_i(x) = 0 & \text{active at } x^* \end{cases}$

In some situations... $\mathcal{X} \neq \emptyset$ but $\mathcal{X}_{\text{opt}} = \emptyset$

Local and global optima (thm) Given $\min_{x \in \mathcal{X}} f_0(x)$ (f_0, \mathcal{X} convex) it holds that: (i) if $x \in \mathcal{X}$ is a local minimizer \rightarrow is a **global minimizer**, (ii) the optimal set \mathcal{X}_{opt} is a convex set.

PROBLEM TRANSFORMATIONS: how to formulate the problem into an "equivalent" way?

(i) Affine variable transformation

Original problem

$$\min_{x \in \mathcal{X}} f_0(x)$$

Transformed problem

$$\min_{x \in \mathcal{X}} \varphi(f_0(x))$$

$\varphi : \mathbb{R} \rightarrow \mathbb{R}$ must be continuous and strictly increasing over \mathcal{X}

(ii) Addition of **slack variables**

Original Problem

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^r \varphi_i(x)$$

$$\text{s.t. } f_i(x) \leq 0, h_i(x) = 0$$

Transformed Problem

$$\min_{x, t} \sum_{i=1}^r t_i$$

$$\text{s.t. } f_i(x) \leq 0, h_i(x) = 0$$

$$\varphi_i(x) \leq t_i$$

This transformation is effective when the objective is the sum of functions φ_i .

(iii) Epigraphic formulation

Original problem

$$p^* = \min_{x \in \mathbb{R}^n} f_0(x)$$

$$\text{s.t. } f_i(x) \leq 0, h_i(x) = 0$$

Transformed problem

$$g^* = \min_{x \in \mathbb{R}^n, t \in \mathbb{R}} t$$

$$\text{s.t. } f_i(x) \leq 0, h_i(x) = 0$$

$$f_0(x) \leq t$$

This type of transformation can be applied for all problems in a way that the objective function becomes linear, pushing the original objective into the constraints.

(iv) Replacement $h_i(x) = 0 \leftrightarrow h_i(x) \leq 0$

Original problem

$$p^* = \min_{x \in \mathbb{R}^n} f_0(x)$$

$$\text{s.t. } f_i(x) \leq 0, h_i(x) = 0$$

Transformed problem

$$g^* = \min_{x \in \mathbb{R}^n} f_0(x)$$

$$\text{s.t. } f_i(x) \leq 0, h_i(x) = 0$$

I can do this substitution if...

f_0 strictly decreasing, $x : h(x) < 0 \in \text{relint}(\mathcal{X}) \leftarrow$ (minimization)

f_0 strictly increasing, $x : h(x) < 0 \in \text{relint}(\mathcal{X}) \leftarrow$ (minimization)

Optimality (Prop) Consider $\min_{x \in \mathcal{X}} f_0(x)$ with f_0, \mathcal{X} convex. A feasible solution $x \in \mathcal{X}$ is **optimal** $\iff \forall y \in \mathcal{X} \nabla f_0(x)^T (y - x) \geq 0$. **Geometric interpretation:** there is no direction for which the objective function can decrease.

Categories of directions There are directions v_{\pm} for which: $\begin{cases} \nabla f_0(x)^T \cdot v_+ > 0 & f_0 \text{ increase} \\ \nabla f_0(x)^T \cdot v_+ < 0 & f_0 \text{ decrease} \rightarrow \text{descent direction} \end{cases}$

Optimality (unconstrained) $x \in \mathcal{X}$ is optimal $\iff \nabla f_0(x) = 0$

Optimality (equality constrained) $x \in \mathcal{X}_{opt} \iff Ax = b, \exists \nu \in \mathcal{R}^m : \nabla f_0(x) + A^T \nu = 0$

Optimality (inequality constrained) Given x feasible, let $\mathcal{A}(x) = \{i : f_i(x) = 0\}, x \in \mathcal{X}_{opt} \iff \exists \lambda_i \geq 0, i \in \mathcal{A}(x) \nabla f_0(x) + \sum_{i \in \mathcal{A}(x)} \lambda_i \nabla f_i(x) = 0$.

8. Quadratic programs (QP)

Quadratic function $f_0(x) = \frac{1}{2}x^T Hx + c^T x + d$ (if $H = 0$, f_0 is linear) The matrix $\begin{cases} H \succeq 0 & f_0 \text{ convex, elliptical paraboloid} \\ H \succ 0 & f_0 \text{ strongly convex} \\ H \preceq 0 & f_0 \text{ concave} \end{cases}$

Unconstrained minimization of f_0

Linear case ($H = 0$)

$$f_0 = c^T x + d, \text{ two cases: } \begin{cases} p^* = -\infty & c \neq 0 \\ p^* = d & c = 0 \end{cases}$$

Quadratic case ($H \neq 0$) $\rightarrow \nabla f_0(x) = 0$ is applied

$\nabla f_0(x) = Hx + c = 0 \iff Hx = -c$. 2 cases:

(i) $c \notin \mathcal{R}(H) \rightarrow p^* = -\infty$ (unbounded below)

(ii) $c \in \mathcal{R}(H) \rightarrow \begin{cases} x^* = -H^\dagger c + \zeta, \zeta \in \mathcal{N}(H) & \text{in general} \\ x^* = -H^{-1}c & \lambda_i > 0, \forall i \end{cases}$

Polyhedron $\mathcal{P} = \{x : a_i^T x \leq b_i, i = 1, \dots, m\} = \{x \in \mathbb{R}^n : Ax \leq b, A \in \mathbb{R}^{m,n}, b \in \mathbb{R}^m\}$ (cvx: intersect. of halfspaces)

...properties: The image of \mathcal{P} through an affine map is still a polyhedron — The set $\mathcal{P} = \{x \in \mathbb{R}^n : Ax \leq b, Cx = d\}$ is a polyhedron. It can be obtained by properly parametrizing x . Polyhedron+Bounded \rightarrow **Polytope**

$$\text{QP (standard form)} \min_{x \in \mathbb{R}^n} \frac{1}{2}x^T Hx + c^T x + d$$

$$\text{subject to: } Ax \leq b, Fx = g$$

The problem is tractable $\iff H \succeq 0$

Example(1): Markowitz's Portfolio

$r_i, x_i \rightarrow$ return and investment for asset a_i

$$\hat{r} = \mathbb{E}\{r\}, \Sigma \doteq \mathbb{E}\{(r - \hat{r})(r - \hat{r})^T\}$$

$$f_0(x) = \underbrace{x^T \Sigma x}_{\text{risk}} - \underbrace{\gamma \hat{r}^T x}_{\text{return}}, \gamma > 0$$

constraints: $x \geq 0, \sum_i x_i = 1$

Example(2): LASSO Problem

ℓ_1 -penalty: $\lambda \|x\|_1 = \lambda \sum_i |x_i|$ (use of slack variables...)

Resulting problem:

$$\min_{x, u} \|Ax - y\|_2^2 + \lambda \sum_{i=1}^n u_i$$

subject to: $x_i \leq u_i, x_i \geq -u_i, i = 1, \dots, n$

9. Linear programs (LP)

Inequality form

$$\begin{aligned} \min_x c^T x \\ \text{s.t. } Ax \leq b, A_{eq}x = b_{eq} \end{aligned}$$

without loss of generality we can assume that $d = 0$: being a constant term, nothing change.

→ problem transformation →

$$\begin{aligned} \text{(slack variables)} \quad e \geq 0, x_+ \geq 0, x_- \geq 0 \\ \tilde{x} = (x_+, x_-, e), \forall x = x_+ - x_- \\ \text{objective } c^T x_+ - c^T x_- \\ \text{constraints } Ax_+ - Ax_- + e = b \\ A_{eq}x_+ - A_{eq}x_- = b_{eq} \\ \tilde{A} = \begin{bmatrix} A & -A & I \\ A_{eq} & -A_{eq} & 0 \end{bmatrix}, \tilde{b} = \begin{bmatrix} b \\ b_{eq} \end{bmatrix} \\ \tilde{c} = [c^T \quad -c^T \quad 0] \end{aligned}$$

Standard form

$$\begin{aligned} \min_x \tilde{c}^T \tilde{x} \\ \text{s.t. } \tilde{A}\tilde{x} = \tilde{b}, \tilde{x} \geq 0 \end{aligned}$$

Can be solved by using:

- (i) SIMPLEX ALGORITHM
- (ii) INTERIOR POINT ALGORITHM

Geometric interpretation The halfspace $\mathcal{H}(x_f) = \{x \in \mathbb{R}^n : c^T(x - x_f) = 0, c \in \mathbb{R}^n\}$ is important $x_f \in \mathcal{X}$ is optimal $\iff c^T x \geq c^T x_f, \forall x \in \mathcal{X} \iff c^T(x - x_f) \geq 0, \forall x \in \mathcal{X}$ - if x_f is not optimal \iff there exists at least one x for which $c^T(x - x_f) < 0$ (feasible descent direction) → the optimal solution is on vertex/edge/facet of the polyhedron representing the feasible set which is totally contained in the halfspace $\mathcal{H}_+(x_f)$

Feasible set and solutions The following situations can arise:

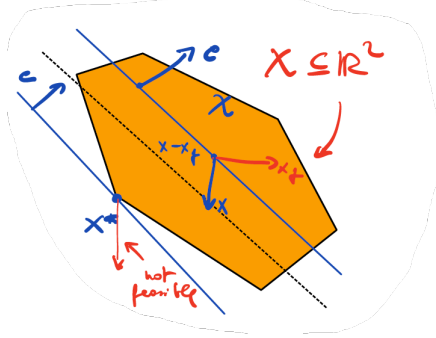
$$\begin{cases} \mathcal{X} = \emptyset & p^* = +\infty \\ \mathcal{X} \text{ bounded} & x^* \text{ is on a vertex/edge/facet} \\ \mathcal{X} \text{ unbounded} & \exists c : p^* = -\infty \text{ unbounded below} \end{cases}$$

Example(1): Diet problem

Most economical diet (total cost: $c^T x$) satisfying the nutritional requirements ($Ax \geq b$).

n foods, their cost is c_i , m nutritional element, their recommended quantity b_i . a_{ij} quantity of the nutrient i in the j -th food. x_i quantity in the diet of the i -th food.

$$\begin{aligned} \min_{x \in \mathbb{R}^n} c^T x \\ \text{s.t. } Ax \geq b, x \geq 0 \end{aligned}$$



Example(2): Max flow problem

Maximize a certain quantity f , sent from a source to a sink, through a network represented by a digraph.

- C (matrix of capacities), c_{ij} (link capacity from i to j)
- X (matrix of effective flow) - clearly $x_{ij} \leq c_{ij}$
- $\phi_{in} = \sum_{i=1}^n x_{ik}, \phi_{out} = \sum_{j=1}^n x_{kj}$

$$\text{Constraints } \begin{cases} \phi_{in} - \phi_{out} = -f & \text{at source (node 1)} \\ \phi_{in} - \phi_{out} = f & \text{at sink} \\ \phi_{in} - \phi_{out} = 0 & \text{intermediate (node } n) \end{cases}$$

$$\text{Objective } \max_{x \in \mathbb{R}^n, f \in \mathbb{R}} f \iff -\min_{x \in \mathbb{R}^n, f \in \mathbb{R}} -f$$

$$-p^* = \min_{x \in \mathbb{R}^n, f \in \mathbb{R}} -f$$

$$\text{s.t. } (X^T - X)\mathbf{1} = \begin{bmatrix} -f \\ 0 \\ f \end{bmatrix}, 0_{n \times n} \leq X \leq C$$

10. Second-order cone programs (SOCP)

Second order cone $\mathcal{K}_p = \{(u, t) : u \in \mathbb{R}^{p-1}, t \in \mathbb{R}, \|u\|_2 \leq t\}$ extension of $\mathcal{K} = \{(x, y, z) : \sqrt{x^2 + y^2} \leq z\}$

\mathbb{R}^n -cone $\mathcal{K}_n = \{x \in \mathbb{R}^n : \sqrt{x_2^2 + x_3^2 + \dots + x_n^2} \leq x_1\}$

Rotated cone $\mathcal{K}_n^r = \{x \in \mathbb{R}^n : x_3^2 + x_4^2 + \dots + x_n^2 \leq 2x_1x_2, x_1, x_2 > 0\}$

Useful equivalence $\|u\|_2^2 \leq yz, y \geq 0, z \geq 0 \iff \left\| \begin{pmatrix} 2u \\ y-z \end{pmatrix} \right\|_2 \leq y+z$

Standard form for SOCP

$$\begin{aligned} \min_x c^T x \\ \text{s.t. } Ax = b, x^i \in \mathcal{K}_{n_i} \end{aligned}$$

SOC constraints

Given $u = Ax + b, t = c^T x + d$

$$\mathcal{K} = \{x : \|Ax + b\|_2 \leq c^T x + d\}$$

I can rewrite the problem as:

$$\begin{aligned} \min_x c^T x \\ \text{s.t. } \|A_i x + d_i\|_2 \leq c_i^T x + b_i \end{aligned}$$

x^i is the i -th block of the decision variable, n_i its dimension.

SOCP class encapsulate also the Linear and Quadratic programs.

LP as SOCP \implies

$$\begin{aligned} \min_x c^T x \\ \text{s.t. } a_i^T x \leq b_i, i = 1, \dots, m \end{aligned}$$

problem transformation

$$\begin{aligned} \min_x c^T x \\ \text{s.t. } \|C_i x + d_i\|_2 \leq b_i - a^T x \\ C_i = 0, d_i = 0, i = 1, \dots, m \end{aligned}$$

QP as SOCP \implies

$$\begin{aligned} \min_x x^T Q x + c^T x + d \\ \text{s.t. } a_i^T x \leq b_i, i = 1, \dots, m \end{aligned}$$

problem transformation

$$\begin{aligned} \min_{x,y} c^T x + y \\ \text{s.t. } \left\| \begin{pmatrix} 2Q^{1/2} \\ y-1 \end{pmatrix} \right\|_2 \leq y+1 \\ \text{s.t. } a_i^T x \leq b_i, i = 1, \dots, m \end{aligned}$$

Sum of Norms (SON) $\min_x \sum_{i=1}^p \|A_i x - b_i\|_2 \longrightarrow \min_{x,y} \sum_{i=1}^p y_i \text{ s.t. } \|A_i x - b_i\|_2 \leq y_i, i = 1, \dots, p, y_i \text{ slack variables}$

Max of Norms $\min_x \max_{i=1, \dots, p} \|A_i x - b_i\|_2 \longrightarrow \min_{x,y} y \text{ s.t. } \|A_i x - b_i\|_2 \leq y, i = 1, \dots, p$

Example(1): Fermat-Weber Point

Where to locate a warehouse in order to serve in the best way some location services?

$x \in \mathbb{R}^2$: position of the warehouse

$y_i \in \mathbb{R}^2, i = 1, \dots, m$: position of the location services

I have to solve: $\min_x \frac{1}{m} \sum_{i=1}^m \|x - y_i\|_2$ (use SON)

$$\begin{aligned} \min_{x,t} \quad & \frac{1}{m} \sum_{i=1}^m t_i \\ \text{s.t.} \quad & \|x - y_i\|_2 \leq t_i, \quad i = 1, \dots, m \end{aligned}$$

SOC Program

11. CVX modeling

Example (LP Program)

```
cvx_begin [quiet]
    cvx_solver mosek
    variable(s) x(2)      % name(dimension)
    minimize (c'*x)        %objective function
    subject to
        A*x <= b;          %constraints
cvx_end
```

Example(2): LP with probability constraints

How to solve an LP problem when one or more data are random or uncertain?

In particular:

- a_i random normally distributed vectors $\begin{cases} \mathbb{E}\{a_i\} = \bar{a}_i \\ \text{var}(a_i) = \Sigma_i \succ 0 \end{cases}$
 - $a_i^T x$ is a random variable $\rightarrow (\mu = \bar{a}_i^T x, \sigma^2 = x^T \Sigma_i x)$,
 - We know a-priori $P\{a_i^T x \leq b_i\} \geq p_i, i = 1, \dots, m$
 If $p_i > 0.5$ the LP problem can be rewritten as:

$$\begin{aligned} \min_x \quad & c^T x \\ \text{s.t.} \quad & \bar{a}_i^T x \leq b_i - \Phi^{-1}(p_i) \|\Sigma_i^{1/2} x\|_2 \end{aligned}$$

SOC Program

You can write the problem also in a non-standard form, cvx will recast the problem in a standard form.

Variables: $\begin{cases} x & \text{the minimizer} \\ \text{cvx.optval} & \text{the value } p^* \\ \text{cvx.status} & \text{Solved, Unfeasible, Unbounded} \end{cases}$

useful function: `quad_form(x,Q)`, returns $x^T Q x$ if $Q \succ 0$

12. General optimization and Lagrangian Duality

Langrangian $\mathcal{L}(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x) - \lambda_i, \nu_i$ **Lagrange multipliers** - $f_0(x) \geq \mathcal{L}(x, \lambda, \nu)$

Lagrangian dual function $g(\lambda, \nu) = \inf_{x \in \mathcal{D}} \mathcal{L}(x, \lambda, \nu), \lambda > 0, g: \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$ - **Properties** $\begin{cases} \text{jointly concave in } (\lambda, \nu) \\ g(\lambda, \nu) \leq p^* \quad \forall \lambda > 0, \forall \nu \end{cases}$

Dual problem (What is the best $g(\lambda, \nu)$)? $d^* = \max_{\lambda, \nu} g(\lambda, \nu)$ subject to: $\lambda > 0 - d^* \leq p^*$ (...is always convex)

Duality gap $\delta^* = p^* - d^* = \begin{cases} = 0 & \text{strong duality (further conditions are needed)} \\ > 0 & \text{weak duality} \end{cases}$

Slater's condition for STRONG DUALITY Let $f_i(x)$ convex functions, $h_i(x)$ affine functions, $f_i(x), i = 1, \dots, k \leq m$ affine, if $\exists x \in \text{relint } \mathcal{D}: f_i(x) \leq 0, i = 1, \dots, k, f_{k+1}(x) < 0, \dots, f_m(x) < 0, h_i(x) = 0, i = 1, \dots, p$ then: $p^* = d^*$ (there is no gap), moreover if the problem is not unbounded below $\exists(\lambda^*, \nu^*): g(\lambda^*, \nu^*) = d^* = p^*$.

Primal problem

$$p^* = \min_x \max_{\lambda, \nu} \mathcal{L}(x, \lambda, \nu)$$

Dual problem

$$d^* = \max_{\lambda, \nu} \min_x \mathcal{L}(x, \lambda, \nu)$$

Primal solution from dual solution (Required: strong duality) $d^* = p^* = f_0(x^*), d^* = g(\lambda^*, \nu^*) = \mathcal{L}(x, \lambda^*, \nu^*)$

$f_0(x^*) = \mathcal{L}(x^*, \lambda^*, \nu^*) = f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) + \sum_{i=1}^p \nu_i^* h_i(x^*) \rightarrow \begin{cases} \sum_{i=1}^m \lambda_i^* f_i(x^*) = 0 & \text{complementary slackness} \\ x^* & \text{minimizer (wrt } x) \text{ of } \mathcal{L}(x, \lambda^*, \nu^*) \end{cases}$

Complementary slackness For the i -th constraint: either $\lambda_i = 0, f_i(x) \leq 0$, or $\lambda_i > 0, f_i(x) = 0$

Solution recovery If f_0 convex, $\nabla_x \mathcal{L}(x, \lambda^*, \nu^*) = 0$ gives a **global minimizer** (unique if f_0 strictly convex).

KKT conditions for optimality (optimal $\iff \delta^* = 0$) $\begin{cases} 1. \text{ primal feasibility} & f_i(x^*) \leq 0, h_i(x^*) = 0 \\ 2. \text{ dual feasibility} & \lambda_i^* \geq 0, i = 1, \dots, m \\ 3. \text{ complementary slackness} & \sum_{i=1}^m \lambda_i^* f_i(x^*) = 0 \\ 4. \text{ Lagrangian stationarity} & \nabla_x \mathcal{L}(x, \lambda^*, \nu^*)|_{x=x^*} = 0 \end{cases}$
 f_0 differentiable, strong duality holds.

Sensitivity of the optimal solution p^*

Perturbed Problem

$$\begin{aligned} p^*(u, v) &= \min_{x \in \mathbb{R}^n} f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq u_i, h_i(x) = v_i \end{aligned}$$

Remark: $p^*(0, 0) = p^* - u_i > 0 \rightarrow$ relaxing, $u_i < 0$ tightening. When the primal problem is convex and $p^*(u, v)$ differentiable:

$$\lambda_i^* = -\frac{\partial p^*(u, v)}{\partial u_i} \Big|_{(u,v)=(0,0)}, \quad \nu_i^* = -\frac{\partial p^*(u, v)}{\partial v_i} \Big|_{(u,v)=(0,0)}$$

Interpretation: (here I use the complementary slackness)

$\lambda_i^* = 0 \Rightarrow f_i(x^*) < 0$ the constraint is inactive, I don't care about perturbations \rightarrow not resource critical!

$\lambda_i^* > 0 \Rightarrow f_i(x^*) = 0$ constraint active (resource critical) $\rightarrow p_{new}^* = p^* - \lambda_i^* u_i$.

In particular: $\begin{cases} u_i < 0 & p^* \text{ increase (worse solution)} \\ u_i > 0 & p^* \text{ can decrease (better solution)} \end{cases}$

Remember that: the Lagrange multipliers are given by the solution of the dual problem.

13. Gradient Algorithm (GA) (unconstrained case)

General structure: $x_{k+1} = x_k + s_k v_k$

$s_k > 0$, $s_k \in \mathbb{R}$ (step-size)

$v_k \in \mathbb{R}^n$ (update/search direction)

For Gradient Algorithm: $\begin{cases} f_0 \text{ differentiable} \\ x_k \in \text{dom}(f_0) \\ v_k \in \mathbb{R}^n \text{ (chosen wrt } \nabla f_0(x)) \end{cases}$

$f_0 : \mathbb{R}^n \rightarrow \mathbb{R}$, $f_0(x_k + s v_k) \simeq f_0(x_k) + s \nabla f_0(x_k)^T v_k$

$t \rightarrow \infty$, $\delta_k = \lim_{s \rightarrow \infty} \frac{f_0(x_k + s v_k) - f_0(x_k)}{s} = \nabla f_0(x_k)^T v_k$

I want a direction for which f_0 decrease (descent). $\forall k$ moving toward $v_k \rightarrow$ go to min.

Steepest descent direction: $v_k = -\frac{\nabla f_0(x)}{\|\nabla f_0(x)\|_2}$

Algorithm 1 Gradient Algorithm

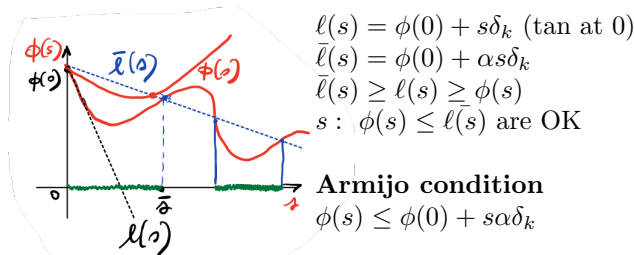
- 1: $k = 0$, choose a descent v_k ($v_k = -\nabla f_0(x_k)$)
- 2: Determine the step-size s_k
- 3: $x_{k+1} = x_k + s_k v_k$
- 4: **if** Stop criterion **then**
- 5: **return** x_k
- 6: **else**
- 7: back to 2
- 8: **end if**

STEPSIZE SELECTION $\phi(s) = f_0(x_k + s v_k)$, $s \geq 0$ (restriction of f_0 along v_k)

$\phi(0) = f_0(x_k)$, I want $s > 0$: $f(x_{k+1}) = \phi(s) < \phi(0)$

Exact line-search (best s possible) (non-convex)

$$s^* = \arg \min_{s \geq 0} \phi(s)$$



Algorithm 2 Backtracking line-search

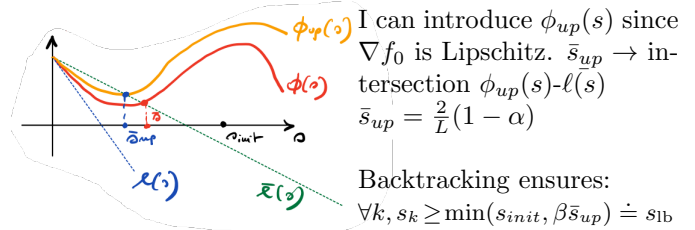
- 1: $\alpha \in (0, 1)$, $\beta \in (0, 1)$, $s_{init} = 1$, v_k descent
- 2: **if** $f_0(x_k + s v_k) \leq f_0(x_k) + s \alpha \nabla f_0(x_k)^T v_k$ **then**
- 3: **return** $s_k \leftarrow s$
- 4: **else**
- 5: $s = \beta s$, back to 2 \triangleright update until not Armijo's meet
- 6: **end if**

CONVERGENCE OF GA: STOPPING CRITERION

$\nabla f_0(x)$, $\exists L : \|\nabla f_0(y) - \nabla f_0(x)\|_2 \leq L \|y - x\|_2, \forall x, y$ (Lipschitz)

$$\Rightarrow f_0(x) \leq f_0(y) + \nabla f_0(x)^T (x - y) + \frac{L}{2} \|x - y\|_2^2$$

strongly cvx, quadratic approximant



It holds that (GA converges to a stationary point):

$$\alpha s_{lb} \sum_{i=0}^k \|\nabla f_0(x_i)\|_2^2 \leq f_0(x_0) - f_0^*, \lim_{k \rightarrow \infty} \|\nabla f_0(x_i)\|_2 = 0$$

Stopping criterion $\rightarrow \|\nabla f_0(x)\|_2 < \varepsilon, \varepsilon > 0$

f_0 convex $\rightarrow x^*$ is a global minimizer.

Convergence $\begin{cases} O(1/\sqrt{k}) & f_0 \text{ generic} \\ O(1/\varepsilon) & f_0 \text{ convex (sublinear)} \\ O(1/\log(1/\varepsilon)) & f_0 \text{ strongly convex (linear)} \end{cases}$

Grad. Algorithm as minimization of $q(x)$

GA can be interpreted also as minimization of the quadratic approximant when $\|x - x_k\|_2$ is small $\rightarrow f_0(x) \simeq q(x) = f_0(x_k) + \nabla f_0(x_k)^T (x - x_k)$

$$\nabla q(x) = \nabla f_0(x_k) + \frac{1}{s} (x - x_k) = 0 \Leftrightarrow x_{k+1} = x_k - s \nabla f_0(x_k)$$

Gradient Algorithm

Conclusion: x is updated $\forall k$ by minimizing $q(x_k)$. \square

14. Newton Algorithm (NA) (unconstrained case)

Key concept: Finding (starting from x_k) the roots of a non linear function $g : \mathbb{R} \rightarrow \mathbb{R}$. We want the roots of $\nabla f_0(x) = 0$.

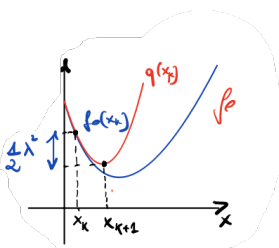
$\tilde{g}(x) = g(x_k) + g'(x_k)(x - x_k)$. x_{k+1} is given by $\tilde{g}(x) = 0$

NEWTON METHOD

$$x_{k+1} = x_k - \frac{g(x_k)}{g'(x_k)} = x_k - (\nabla^2 f_0(x_k))^{-1} \nabla f_0(x_k) \quad (4)$$

$$q_0(x) = f(x_k) + \nabla f_0(x_k)^T (x - x_k) + \frac{1}{2} (x - x_k)^T \nabla^2 f_0(x_k) (x - x_k)$$

quadratic approximant of f_0



$$\begin{aligned} \arg \min_x q_0(x) &= x_{k+1} \\ \min_x q_0(x) &= f(x_{k+1}) \\ \min_x q_0(x) - f(x) &= \frac{1}{2} \lambda^2(x) \\ \lambda^2 &= \nabla f_0^T \nabla^2 f_0^{-1} \nabla f_0 \end{aligned}$$

Newton decrement
 where $\nabla f_0 \leftrightarrow \nabla f_0(x_k)$

Damped Newton Method $x_{k+1} = x_k + t v$

$$v = -\nabla^2 f_0(x_k)^{-1} \nabla f_0(x_k), \nabla f_0(x_k)^T \cdot v = -\lambda^2 < 0 \text{ (descent)}$$

Algorithm 3 Choice of t (step-size)

- 1: $\alpha \in (0, \frac{1}{2}]$, $\beta \in (0, 1)$, $t = 1$
- 2: **while** $f_0(x_k + t v) > f_0(x_k) + \alpha t \nabla f_0(x_k)^T v$ **do** $t \leftarrow t \beta$
- 3: **end while**
- 4: $x_{k+1} = x_k + t v$

Convergence properties f_0 strongly convex, $\nabla f_0, \nabla^2 f_0$ Lipschitz continuous, $\eta \in (0, m^2/L)$.

2 phases: $\begin{cases} \|\nabla f_0(x_k)\| \geq \eta & \text{Damped phase} \\ \|\nabla f_0(x_k)\| < \eta & \text{Quadratically convergent phase} \end{cases}$

Stopping criterion $f_0(x_k) - f_0^* \leq \lambda_k^2$

OTHER ASPECTS (1): EQUALITY CONSTRAINTS

I choose $x_0 : A x_0 = b$, and $x_{k+1} = x_k + t v$, where

$$v = \arg \min_{z \in \mathcal{N}(A)} \nabla f_0(x)^T (z - x) + \frac{1}{2} (z - x)^T \nabla^2 f_0(x_k) (z - x)$$

In this way, since $A x_{k+1} = A x_k + t A v = b$, x_k remains in the feasible set, $A v = 0$.

OTHER ASPECTS(2): QUASI-NEWTON METHODS

Secant condition $H(\nabla f(x) - \nabla f(y)) = (x - y) \Rightarrow$ a matrix H satisfying such a condition can approximate the Hessian.

Algorithm 4 Quasi-Newton methods

- 1: $H_k = I_n$
- 2: Update s.t.: $H_{k+1} \nabla f_0(x_{k+1}) - \nabla f_0(x_k) = x_{k+1} - x_k$

15. Approaches for constrained optimization

When there are constraints solving $\nabla f_0(x) = 0$ it is not sufficient, x^* must be feasible and $\mathcal{X} \neq \mathbb{R}^n$

1st approach: *Projected Gradient Method*

$\mathcal{P}_{\mathcal{X}}(x) \leftarrow$ projection of x on the feasible set \mathcal{X} (convex, non-empty).

Algorithm 5 Projected Gradient Method

- 1: $k = 0$
- 2: $w_{k+1} = x_k - s_k \nabla f_0(x_k)$ \triangleright Gradient Step
- 3: $x_{k+1} = \mathcal{P}_{\mathcal{X}}(w_{k+1})$ \triangleright ensure feasibility

(used only for set for which is simple to compute projections, otherwise other more general methods are used).

The s_k (step-size) is chosen as follows: $s_k = \bar{s}2^{-t(k)}$

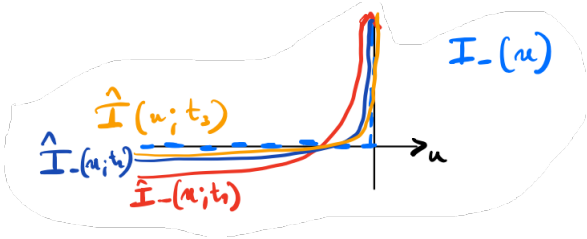
Algorithm 6 Stepsize selection

- 1: $j = 0$
- 2: $z_j = \mathcal{P}_{\mathcal{X}}(x_k - \bar{s}2^{-j} \nabla f_0(x_k))$
- 3: **if** $f(z_j) \leq f(x_k) - \alpha \nabla f_0(x_k)^T (x_k - z_j)$ **then**
- 4: $t(k) = j$
- 5: **else**
- 6: goto 2
- 7: **end if**

2nd approach: *Barrier Method*

Here we want to solve (f_0, \dots, f_m convex and smooth):

$$\begin{aligned} p^* &= \min_x f_0(x) \\ \text{s.t. } f_i(x) &\leq 0, \quad i = 1, \dots, m \\ Ax &= b \end{aligned}$$



The problem can be rewritten as:

$$p^* = \min_x f_0(x) + \sum_{i=1}^m I_-(f_i(x)) \quad (5)$$

$$I_-(u) = \begin{cases} 0 & u \leq 0 \\ \infty & u > 0 \end{cases} \quad (\text{Indicator function}) \quad (6)$$

$I_-(u)$ non-differentiable $\rightarrow \hat{I}_-(u; t) = -\frac{1}{t} \log(-u)$
 $\phi(t) \rightarrow \text{logarithmic barrier}$

$$p^*(t) = \min_x f_0(x) - \frac{1}{t} \sum_{i=1}^m \log(-f_i(x)) = \quad (7)$$

$$= \min_x t f_0(x) + \phi(t) \quad (8)$$

$$\text{s.t. } Ax = b$$

Parametrizing the x , (8) becomes unconstrained (GA, NM can be used).

For $t \rightarrow \infty$ (8) \rightarrow (5) that is $p^*(t) \rightarrow p^*$, $x^*(t) \rightarrow x^*$

Central path $\{x^*(t) : t > 0\}$, $p^*(t)$ ε -suboptimal $\varepsilon = m/t$

Algorithm 7 Sequential Barrier Method

- 1: x_0 (strictly feasible), $t \leftarrow t_0$, $\mu > 1$, $\varepsilon > 0$
- 2: **loop**
- 3: Solve $\min_{Ax=b} t f_0(t) + \phi(t)$ \triangleright Centering step
- 4: $x \leftarrow x^*(t)$
- 5: **if** $m/t < \varepsilon$ **then** \triangleright Stopping criterion
- 6: **break**
- 7: **else**
- 8: $t \leftarrow \mu t$ \triangleright Increase t (better approx of $I_-(u)$)
- 9: **end if**
- 10: **end loop**

Phase I Problem: starting from a feasible x_0

$$\begin{aligned} \tilde{s} &= \min_{x,s} s && \text{Solved by using the barrier method, starting from:} \\ &\text{s.t. } f_i(x) \leq s, && x_0 : Ax_0 = b, \\ &Ax = b && s_0 = \eta + \max_i f_i(x), \quad \eta > 0 \end{aligned}$$

$\rightarrow \tilde{s} < 0$. Since $f_i(\tilde{x}) \leq \tilde{s} \leq 0 \rightarrow x_0 = \tilde{x}$

$\rightarrow \tilde{s} > 0$. $\nexists x : f_i(x) \leq 0 \rightarrow$ problem infeasible

$\rightarrow \tilde{s} = 0$. Only of theoretical interest

(9) **Remark:** For carrying out the *centering step* the NEWTON METHOD tailored for equality constraints is used.

16. Geometric Programs(GP)

For GP: **variables** \rightarrow positive (physical quantities), **objective/constraints** \rightarrow non-negative linear combination of positive monomials.

Positive monomial Given $x \in \mathbb{R}^n$, $a \in \mathbb{R}^n$, $c > 0$, $x > 0$ $cx^a = cx_1^{a_1} x_2^{a_2} \dots x_n^{a_n} \in \mathbb{R}^+$

Posynomial $f : \mathbb{R}_{++}^n \rightarrow \mathbb{R}$, $f(x) = \sum_{i=1}^k c_i x^{a(i)}$, $c_i, x > 0$, $a(i) \in \mathbb{R}^n$ - **Generalized posynomial** $\left\{ \begin{array}{l} \text{pointwise maximum} \\ \text{fractional power} \\ \text{addition/multiplication} \end{array} \right.$
 composite fun. of posynomials

Monomials, Posynomials and Generalized posynomials **are not convex!** \rightarrow Problem transformation is needed

Convex form for monomials $y_i = \log x_i \rightarrow \tilde{g}(x) = cx_1^{a_1} \dots x_n^{a_n} = e^{\log c} e^{\log x_1^{a_1}} \dots e^{\log x_n^{a_n}} = e^{a_1 y_1 + \dots + a_n y_n + \log c} = e^{a^T y + b}$. I can take the log (f convex \circ f increasing) obtaining $g(x) = \log \tilde{g}(x) = a^T y + b \rightarrow$ Linear Program

Convex form for posynomials $\tilde{f}(x) = \sum_{i=1}^k c_i x^{a(i)} = \sum_{i=1}^k e^{a(i)^T y + b_i} \rightarrow \tilde{y}(x) = \log \left(\sum_{i=1}^k e^{a(i)^T y + b_i} \right) = \text{lse}(Ax + b)$

log-sum-exp (lse) function is convex - $A \in \mathbb{R}^{k,n}$, $b = [b_1 \dots b_n]^T$

Standard form for GP Here $f_0(x), f_i(x)$ are posynomials, $h_i(x)$ are (possibly) monomials

Geometric programs in standard form

$$\begin{aligned}
 \min_x \quad & \sum_{k=1}^{K_0} c_k x^{a_{(k)}} \\
 \text{s.t.} \quad & \sum_{k=1}^{K_i} c_k x^{a_{(k)}} \leq 1, \quad i = 1, \dots, m \\
 & g_i x^{r_{(i)}} = 0, \quad i = 1, \dots, p
 \end{aligned}
 \qquad
 \begin{aligned}
 \min_y \quad & \text{lse}(A_0 y + b_0) \\
 \text{s.t.} \quad & \text{lse}(A_i y + b_i) \leq 0 \\
 & Ry + h = 0 \\
 & A_0 \in \mathbb{R}^{K_0, n}, b_0 \in \mathbb{R}^{K_0}, A_i \in \mathbb{R}^{K_i, n}, b_i \in \mathbb{R}^{K_i}
 \end{aligned}$$

Generalized GP f_0, \dots, f_m are generalized posynomials \rightarrow GGP (Generalized GP)

Fractional power $f_1(x)^\alpha + f_2(x)^\beta \leq 1 \Rightarrow \underbrace{f_1(x) \leq t_1, f_2(x) \leq t_2, t_1^\alpha + t_2^\beta \leq 1}_{\text{GP constraints}}$

Pointwise maximum power $\max(f_1(x), f_2(x)) + f_3(x) \leq 1 \Rightarrow \underbrace{f_1(x) \leq t, f_2(x) \leq t, t + f_3(x) \leq 1}_{\text{GP constraints}}$

APPENDIX

Schur complement $M = \begin{bmatrix} A & X^T \\ X & B \end{bmatrix}, A, B \in \mathbb{S}^n, M \succeq 0 \iff S = A - XB^{-1}X^T \succeq 0$

Young inequality If $a \geq 0, b \geq 0, p > 1, q > 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$ then $ab \leq \frac{1}{p}a^p + \frac{1}{q}b^q$ (Jensen's inequality can be used in order to demonstrate it)

Maxima of inner product over a ball $\max_{\|x\|_p \leq 1} x^T y \begin{cases} \max_{\|x\|_2 \leq 1} x^T y = \|y\|_2, \quad x^* = \frac{y}{\|y\|_2} & p = 2 \\ \max_{\|x\|_\infty \leq 1} x^T y = \sum_{i=1}^n |y_i| = \|y\|_1, \quad x_\infty^* = \text{sgn}(y) & p = \infty \\ \max_{\|x\|_1 \leq 1} x^T y = \max_i |y_i| = \|y\|_\infty & p = 1 \\ [x_1^*]_i = \begin{cases} \text{sgn}(y_i) & \text{if } i = m \\ 0 & \text{otherwise} \end{cases}, \quad i = 1, \dots, n & p = 1 \end{cases}$