Numerical Linear Algebra

Maya Neytcheva, TDB, February-March 2021

Shortly on projectors

Projectors and properties

Definitions:
Consider C^n and a mapping $P : C^n \to C^n$.
P is called a *projector* if $P^2 = P$ (i.e. P is idempotent).
If P is a projector, then $I - P$ is also such:

$$(I - P)^2 = I - 2P + P^2 = I - P.$$

$\mathcal{N}(P) = \{\mathbf{x} \in C^n : P\mathbf{x} = 0\}$ (null space (kernel) of P)
$\mathcal{R}(P) = \{P\mathbf{x} : \mathbf{x} \in C^n\}$ (range of P).

A subspace S is called *invariant under a square matrix A* whenever $AS \in S$.

Properties: I

P1: $\mathcal{N}(P) \bigcap \mathcal{R}(P) = \{0\}$. Indeed,

$$\text{if } \mathbf{x} \in \mathcal{R}(P) \Rightarrow \exists \mathbf{y} : \mathbf{x} = P\mathbf{y} \Rightarrow P\mathbf{x} = P^2\mathbf{y} = P\mathbf{y}$$

$$\Rightarrow \mathbf{y} = \mathbf{x} \Rightarrow \mathbf{x} = P\mathbf{x}.$$

If $\mathbf{x} \in \mathcal{N}(P) \Rightarrow P\mathbf{x} = 0 \Rightarrow \mathbf{x} = P\mathbf{x} \Rightarrow \mathbf{x} = 0$.

Properties: II

P2: $\mathcal{N}(P) = \mathcal{R}(I - P)$

$\mathbf{x} \in \mathcal{N}(P) \Rightarrow P\mathbf{x} = 0$. Then $\mathbf{x} = I\mathbf{x} - P\mathbf{x} = (I - P)\mathbf{x}$.

$\mathbf{x} \in \mathcal{R}(I - P) \Rightarrow \mathbf{x} = (I - P)\mathbf{z} \Rightarrow P\mathbf{x} = P\mathbf{z} - P^2\mathbf{z} = 0 \Rightarrow P\mathbf{x} = 0$.

P3: $C^n = \mathcal{R}(P) \bigoplus \mathcal{N}(P)$.

Properties: III

P4: Given two subspaces K and L of same dimension m, the following two conditions are mathematically equivalent:

(i) No nonzero vector in K is orthogonal to L

(ii) $\forall \mathbf{x} \in C^n \exists$ unique vector $\mathbf{y} : \mathbf{y} \in K, \mathbf{x} - \mathbf{y} \in L^\perp$.

Proof.

(i)⇒(ii): $K \bigcap L^\perp = \{\emptyset\} \Rightarrow C^n = K \bigoplus L^\perp \Rightarrow \forall \mathbf{x} \in C^n : \mathbf{x} = \mathbf{y} + \mathbf{z}$, where $\mathbf{y} \in K$ and $\mathbf{z} \in L^\perp$. Thus, $\mathbf{z} = \mathbf{x} - \mathbf{y} \Rightarrow$ (ii). □

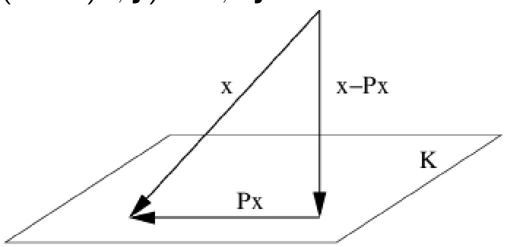
Properties: IV

P5: *Orthogonal* and *oblique* projectors

P is *orthogonal* if $\mathcal{N}(P) = \mathcal{R}(P)^\perp$. Otherwise P is *oblique*.

Thus, if P is orthogonal onto K, then $P\mathbf{x} \in K$ and $(I - P)\mathbf{x} \perp K$.

Equivalently, $((I - P)\mathbf{x}, \mathbf{y}) = 0, \forall \mathbf{y} \in K$.



Properties: V

P6: If P is orthogonal, then $\|P\| = 1$.

Proof.

$\mathbf{x} = P\mathbf{x} + (I - P)\mathbf{x} = \mathbf{y} - \mathbf{z}$.

Then $(\mathbf{y}, \mathbf{z}) = 0 : (P\mathbf{x}, (I - P)\mathbf{x}) = (P\mathbf{x}, \mathbf{x}) - (P\mathbf{x}, P\mathbf{x}) = (P\mathbf{x}, \mathbf{x}) - (P\mathbf{x}, \mathbf{x}) = 0$.

$\Rightarrow \|\mathbf{x}\|_2^2 = \|P\mathbf{x}\|_2^2 + \|(I - P)\mathbf{x}\|_2^2$

$\Rightarrow \|\mathbf{x}\|_2^2 \geq \|P\mathbf{x}\|_2^2 \Rightarrow \frac{\|P\mathbf{x}\|_2^2}{\|\mathbf{x}\|_2^2} \leq 1, \forall \mathbf{x} \in C^n$.

However, for $\widetilde{\mathbf{x}} \in \mathcal{R}(P)$ there holds $\frac{\|P\widetilde{\mathbf{x}}\|_2^2}{\|\widetilde{\mathbf{x}}\|_2^2} = 1$. Thus, $\|P\| = 1$. □

Properties: VI

P7: Any orthogonal projector has only two eigenvalues 0 and 1. Any vector from $\mathcal{R}(P)$ is an eigenvector to $\lambda = 1$. Any vector from $\mathcal{N}(P)$ is an eigenvector to $\lambda = 0$.

Theorem

Let P be orthogonal onto K. Then for any vector $\mathbf{x} \in C^n$ there holds

$$\min_{\mathbf{y} \in K} \|\mathbf{x} - \mathbf{y}\|_2 = \|\mathbf{x} - P\mathbf{x}\|_2. \quad (1)$$

Properties: VII

Proof.

For any $\mathbf{y} \in K$, $P\mathbf{x} - \mathbf{y} \in K$, $P\mathbf{x} \in K$, $(I - P)\mathbf{x} \perp K$
$\|\mathbf{x} - \mathbf{y}\|_2^2 = \|(\mathbf{x} - P\mathbf{x}) + (P\mathbf{x} - \mathbf{y})\|_2^2 =$
$\|\mathbf{x} - P\mathbf{x}\|_2^2 + \|P\mathbf{x} - \mathbf{y}\|_2^2 + 2(\mathbf{x} - P\mathbf{x}, P\mathbf{x} - \mathbf{y}) = \|\mathbf{x} - P\mathbf{x}\|_2^2 + \|P\mathbf{x} - \mathbf{y}\|_2^2$.
Therefore, $\|\mathbf{x} - \mathbf{y}\|_2^2 \geq \|\mathbf{x} - P\mathbf{x}\|_2^2$ $\forall \mathbf{y} \in K$ and the minimum is reached for $\mathbf{y} = P\mathbf{x}$. □

Corollary

Let $K \subset C^n$ and $\mathbf{x} \in C^n$ be given. Then $\min_{\mathbf{y}} \|\mathbf{x} - \mathbf{y}\|_2 = \|\mathbf{x} - \mathbf{y}^*\|_2$ is equivalent to $\mathbf{y}^* \in K$ and $\mathbf{x} - \mathbf{y}^* \perp K$.

Iterative solution methods

- ⇛ Steepest descent
- ⇛ Conjugate gradient method (CG)
- ⇛ Generalized conjugate gradient method (GCG)
- ⇛ ORTHOMIN
- ⇛ Minimal residual method (MINRES)
- ⇛ Generalized minimal residual method (GMRES)
- ⇛ Lanczos method
- ⇛ Arnoldi method
- ⇛ Orthogonal residual method (ORTHORES)

Iterative solution methods

- ⇛ Full orthogonalization method (FOM)
- ⇛ Incomplete orthogonalization method (IOM)
- ⇛ SYMMLQ - variant of CG for symmetric indefinite systems
- ⇛ Biconjugate gradient method (BiCG)
- ⇛ BiCGStab
- ⇛ Conjugate gradients squared (CGS)
- ⇛ Minimal residual method (MR)
- ⇛ Quasiminimal residual method (QMR)
- ⇛ ...

Projection-based iterative methods

General framework – projection methods

Want to solve $\boxed{\mathbf{b} - A\mathbf{x}b = \mathbf{0}}$, $\mathbf{b}, \mathbf{x} \in R^n, A \in R^{n \times n}$.

Instead, choose two subspaces $L \subset R^n$ and $K \subset R^n$ and

$$\boxed{* \text{ find } \widetilde{\mathbf{x}} \in \mathbf{x}^{(0)} + \delta,\ \delta \in K, \text{ such that } \mathbf{b} - A\widetilde{\mathbf{x}} \perp L}$$

K - search space
L - subspace of constraints
$*$ - basic projection step

The framework is known as Petrov-Galerkin conditions.

There are two major classes of projection methods:

- orthogonal - if $K \equiv L$,
- oblique - if $K \neq L$.

Notations:
$\widetilde{\mathbf{x}} = \mathbf{x}^0 + \delta$ - (δ - correction)
$\mathbf{r}^0 = \mathbf{b} - A\mathbf{x}^0$ ($\mathbf{r}^0$ - residual)

$$\boxed{* \text{ find } \delta \in K, \text{ such that } \mathbf{r}^0 - A\delta \perp L}$$

Matrix formulation

Choose a basis in K and L: $V = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m\}$ and $W = \{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_m\}$.
Then, $\widetilde{\mathbf{x}} = \mathbf{x}^0 + \delta = \mathbf{x}^0 + V\mathbf{y}$ for some $\mathbf{y} \in R^m$.

The orthogonality condition can be written as

$$(**)\ \boxed{W^T(\mathbf{r}^0 - AV\mathbf{y})}$$

which is exactly the Petrov-Galerkin condition.
From $(**)$ we get

$$W^T\mathbf{r}^0 = W^TAV\mathbf{y}$$
$$\mathbf{y} = (W^TAV)^{-1}W^T\mathbf{r}^0$$
$$\widetilde{\mathbf{x}} = \mathbf{x}^0 + V(W^TAV)^{-1}W^T\mathbf{r}^0$$

In practice, $m < n$, even $m \ll n$, for instance, $m = 1$.

Matrix formulation, cont.

$$\widetilde{\mathbf{x}} = \mathbf{x}^0 + V(W^TAV)^{-1}W^T\mathbf{r}^0$$

The matrix W^TAV will be small and, hopefully, with a nice structure.

!!! W^TAV should be invertible.

A prototype projection-based iterative method:

Given $\mathbf{x}^{(0)}$; $\mathbf{x} = \mathbf{x}^{(0)}$
Until convergence do:
Choose K and L
Choose basis V in K and W in L
Compute $\mathbf{r} = \mathbf{b} - A\mathbf{x}$
$\mathbf{y} = (W^TAV)^{-1}W^T\mathbf{r}$
$\mathbf{x} = \mathbf{x} + V\mathbf{y}$

Degrees of freedom: m, K, L, V, W.
Clearly, if $K \equiv L$, then $V = W$.

Plan:
(1) Consider two important cases: $L = K$ and $L = AK$
(2) Make a special choice of K.

Property 1:

Theorem

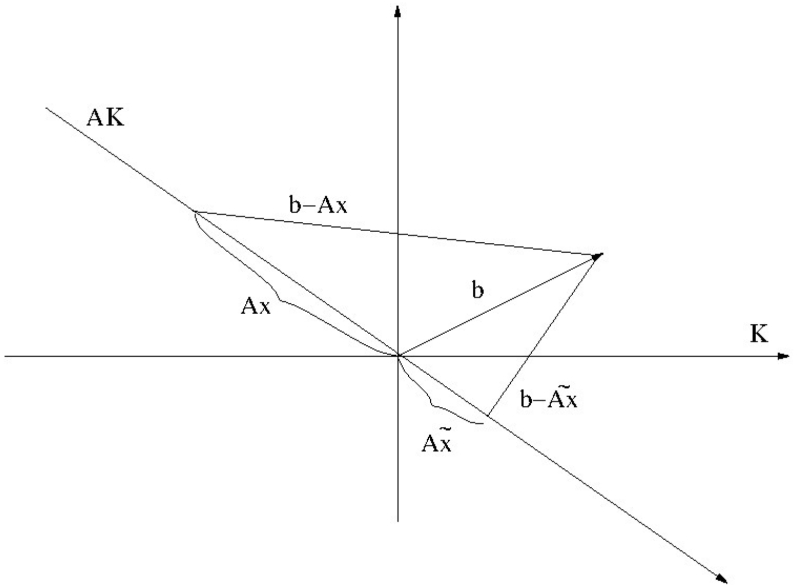
Let A be square, $L = AK$. Then a vector $\widetilde{\mathbf{x}}$ is an oblique projection on K orthogonally to AK with a starting vector $\mathbf{x}^0$ if and only if $\widetilde{\mathbf{x}}$ minimizes the 2-norm of the residual over $\mathbf{x}^0 + K$, i.e.,

$$\|\mathbf{b} - A\widetilde{\mathbf{x}}\|_2 = \min_{\mathbf{x}\in\mathbf{x}^0+K} \|\mathbf{b} - A\mathbf{x}\|_2. \quad (2)$$

Thus, the residual decreases monotonically.

Referred to as *minimal residual methods*
CR, GCG, GMRES, ORTHOMIN

Property 1:



Property 2:

Theorem

Let A be symmetric positive definite, i.e., it defines a scalar product $(A\cdot,\cdot)$ and a norm $\|\cdot\|_A$. Let $L = K$, i.e., $\mathbf{r}^0 - A\widetilde{\mathbf{x}} \perp K$. Then a vector $\widetilde{\mathbf{x}}$ is an orthogonal projection onto K with a starting vector $\mathbf{x}^0$ if and only if it minimizes the A-norm of the error $\mathbf{e} = \mathbf{x}^* - \mathbf{x}$ over $\mathbf{x}^0 + K$, i.e.,

$$\|\mathbf{x}^* - \widetilde{\mathbf{x}}\|_A = \min_{\mathbf{x}\in\mathbf{x}^0+K} \|\mathbf{x}^* - \mathbf{x}\|_A. \quad (3)$$

The error decreases monotonically in the A-norm.
Error-projection methods.

Example: $m = 1$

Consider two vectors: **d** and **e**. Let $K = span\{\mathbf{d}\}$ and $L = span\{\mathbf{e}\}$.
Then $\widetilde{\mathbf{x}} = \mathbf{x}^0 + \alpha\mathbf{d}$ ($\delta = \alpha\mathbf{d}$) and the orthogonality condition reads as:

$$\mathbf{r}^0 - A\delta \perp \mathbf{e} \Rightarrow (\mathbf{r}^0 - A\delta, \mathbf{e}) = 0 \Rightarrow \alpha(A\mathbf{d},\mathbf{e}) = (\mathbf{r}^0,\mathbf{e}) \Rightarrow \alpha = \frac{(\mathbf{r}^0,\mathbf{e})}{(A\mathbf{d},\mathbf{e})}.$$

If $\mathbf{d} = \mathbf{e}$ - the Steepest Descent method (minimization on a line).
If we minimize over a plane - ORTHOMIN.

Choice of K:

$$K = \mathcal{K}^m(A,\mathbf{v}) = \{\mathbf{v}, A\mathbf{v}, A^2\mathbf{v}, \cdots, A^{m-1}\mathbf{v}\}$$

Krylov subspace methods

- $L = K = \mathcal{K}^m(A,\mathbf{r}^0)$ and A spd $\Rightarrow$ CG
- $L = AK = A\mathcal{K}^m(A,\mathbf{r}^0)$ $\Rightarrow$ GMRES

A question to answer:
Why are Krylov subspaces of interest?

How to construct a basis for $\mathcal{K}$?

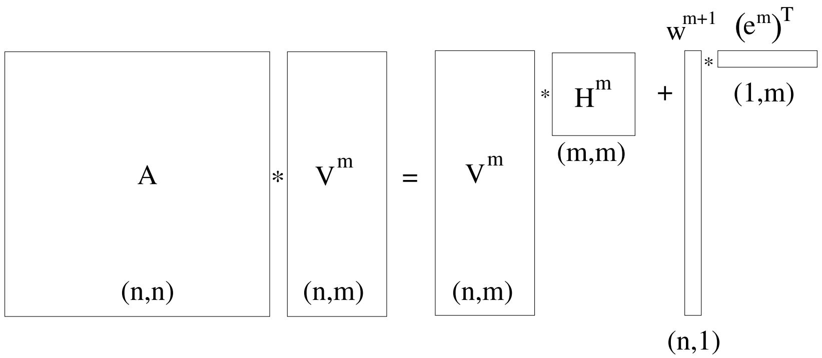
Arnoldi's method for general matrices

Consider $\mathcal{K}^m(A, \mathbf{v}) = \{\mathbf{v}, A\mathbf{v}, A^2\mathbf{v}, \cdots, A^{m-1}\mathbf{v}\}$, generated by some matrix A and vector $\mathbf{v}$.

1. Choose a vector $\mathbf{v}_1$ such that $\|\mathbf{v}_1\| = 1$
2. For $j = 1, 2, \cdots, m$
3. For $i = 1, 2, \cdots, j$
4. $h_{ij} = (A\mathbf{v}_j, \mathbf{v}_i)$
5. End
6. $\mathbf{w}_j = A\mathbf{v}_j - \sum_{i=1}^{j} h_{ij}\mathbf{v}_i$
7. $h_{j+1,j} = \|\mathbf{w}_j\|$
8. If $h_{j+1,j} = 0$, stop
9. $\mathbf{v}_{j+1} = \mathbf{w}_j / h_{j+1,j}$
10. End

The result of Arnoldi's process

- $V^m = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m\}$ is an orthonormal basis in $\mathcal{K}^m(A, \mathbf{v})$
- $AV^m = V^m H^m + \mathbf{w}_{m+1}\mathbf{e}_m^T$



Arnoldi's process - example

$$H^3 = \begin{bmatrix} (A\mathbf{v}_1, \mathbf{v}_1) & (A\mathbf{v}_2, \mathbf{v}_1) & (A\mathbf{v}_3, \mathbf{v}_1) \\ \|\mathbf{w}_1\| & (A\mathbf{v}_2, \mathbf{v}_2) & (A\mathbf{v}_3, \mathbf{v}_2) \\ 0 & \|\mathbf{w}_2\| & (A\mathbf{v}_3, \mathbf{v}_3) \end{bmatrix}$$

Since $V^{m+1} \perp \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m\}$ then it follows that $(V^m)^T A V^m = H^m$.
H^m is an upper-Hessenberg matrix.

Arnoldi's method for symmetric matrices

Let now A be real symmetric matrix. Then the Arnoldi method reduces to the Lanczos method.

Recall: $H^m = (V^m)^T A V^m$
If A is symmetric, then H^m must be symmetric too, i.e., H^m is three-diagonal

$$H^m = \begin{bmatrix} \gamma_1 & \beta_2 & & \\ \beta_2 & \gamma_2 & \beta_3 & \\ & & \ddots & \\ & & \beta_m & \gamma_m \end{bmatrix}$$

Thus, the vectors $\mathbf{v}^j$ satisfy a three-term recursion:

$$\beta_{i+1}\mathbf{v}^{i+1} = A\mathbf{v}^i - \gamma_i \mathbf{v}^i - \beta_i \mathbf{v}^{i-1}$$

Arnoldi	Lanczos
$\mathbf{w}^{(0)}$, $\beta = \lVert\mathbf{w}^{(0)}\rVert$, $\mathbf{v}^{(1)} = \mathbf{w}^{(0)}/\beta$	$\mathbf{w}^{(0)}$, $\beta = \lVert\mathbf{w}^{(0)}\rVert$, $\mathbf{v}^{(1)} = \mathbf{w}^{(0)}/\beta$
For $k = 1, 2, \cdots, m$	Set $\beta_1 = 0$, $\mathbf{v}^{(0)} = \mathbf{0}$
For $i = 1, 2, \cdots, k$	For $k = 1, 2, \cdots, m$
$h_{ik} = (A\mathbf{v}^{(k)}, \mathbf{v}^{(i)})$	$\mathbf{w}^{(k)} = A\mathbf{v}^{(k)} - \beta_k \mathbf{v}^{(k-1)}$
End	$\gamma_k = (\mathbf{w}^{(k)}, \mathbf{v}^{(k)})$
$\mathbf{w}^{(k)} = A\mathbf{v}^{(k)} - \sum_{i=1}^{(k)} h_{ik}\mathbf{v}^{(i)}$	$\mathbf{w}^{(k)} = \mathbf{w}^{(k)} - \gamma_k \mathbf{v}^{(k)}$
$h_{k+1,k} = \lVert\mathbf{w}^{(k)}\rVert$	$\beta_{k+1} = \lVert\mathbf{w}^{(k)}\rVert$
If $h_{k+1,k} = 0$, stop	if $\beta_{k+1} = 0$, stop
$\mathbf{v}^{(k+1)} = \mathbf{w}^{(k)}/h_{k+1,k}$	$\mathbf{v}^{(k+1)} = \mathbf{w}^{(k)}/\beta_{k+1}$
End	End
	Set $T_m = tridiag\{\beta_k, \gamma_k, \beta_{k+1}\}$

Lanczos algorithm to solve symmetric linear systems

Given:	$\mathbf{x}^{(0)}$
Compute	$\mathbf{r}^{(0)} = \mathbf{b} - A\mathbf{x}^{(0)}$, $\beta = \lVert\mathbf{r}^{(0)}\rVert$, $\mathbf{v}^1 = \mathbf{r}^{(0)}/\beta$
Set	$\beta_1 = 0$ and $\mathbf{v}^0 = \mathbf{0}$
For	$j = 1 : m$
	$\mathbf{w}^j = A\mathbf{v}^j - \beta_j \mathbf{v}^{j-1}$
	$\gamma_j = (\mathbf{w}^j, \mathbf{v}^j)$
	$\mathbf{w}^j = \mathbf{w}^j - \gamma_j \mathbf{v}^j$
	$\beta_{j+1} = \lVert\mathbf{w}^j\rVert_2$, if $\beta_{j+1} = 0$, go out of the loop
	$\mathbf{v}^{j+1} = \mathbf{w}^j/\beta_{j+1}$
End	
Set	$T_m = tridiag\{\beta_i, \gamma_i, \beta_{i+1}\}$
Compute	$\mathbf{y}^m = T_m^{-1}(\beta\mathbf{e}_1)$
	$\mathbf{x}^m = \mathbf{x}^0 + V^m \mathbf{y}^m$

Leads to three-term CG.

Direct Lanczos: the factorization of T_m

The coefficients on the direct Lanczos algorithm correspond to the following factorization of T_m:

$$T_m = \begin{bmatrix} \gamma_1 & \beta_2 & & \\ \beta_2 & \gamma_2 & \beta_3 & \\ & & \ddots & \\ & & \beta_m & \gamma_m \end{bmatrix} = \begin{bmatrix} 1 & & & \\ \lambda_2 & 1 & & \\ & & \ddots & \\ & & \lambda_m & 1 \end{bmatrix} \begin{bmatrix} \eta_1 & \beta_2 & & \\ & \eta_2 & \beta_3 & \\ & & \ddots & \\ & & & \eta_m \end{bmatrix}$$

where

i	λ_i	η_i
$i = 1$		$\eta_1 = \gamma_1$
$i = 2, \cdots m$	$\lambda_i = \beta_i/\eta_{i-1}$	$\eta_i = \gamma_i - \lambda_i\beta_{i-1}$

Direct Lanczos

Instead of factorizing at the end, Gauss factorization without pivoting can be performed while constructing T.
Recall $\mathbf{x}^m = \mathbf{x}^0 + V^m L^{-T} L^{-1} \beta \mathbf{e}_1$ and let $G = V^m L^{-T}$ and $\mathbf{z} = L^{-1}\beta\mathbf{e}_1$

Compute	$\mathbf{r}^{(0)} = \mathbf{b} - A\mathbf{x}^{(0)}$, $\xi_1 = \beta = \|\mathbf{r}^{(0)}\|$, $\mathbf{v}^1 = 1/\beta\, \mathbf{r}^{(0)}$
Set	$\lambda_1 = 1, \beta_1 = 1, \mathbf{g}^0 = \mathbf{0}, \beta_1 = 0$ and $\mathbf{v}^0 = \mathbf{0}$
For	$j = 1, 2, \cdots$ until convergence
	$\mathbf{w} = A\mathbf{v}^j - \beta_j \mathbf{v}^{j-1}$
	$\gamma_j = (\mathbf{w}, \mathbf{v}^j)$
	if $j > 1, \lambda_j = \beta_j/\eta_{j-1},\ \xi_j = -\lambda_j \xi_{j-1}$
	$\eta_j = \gamma_j - \lambda_j \beta_j$
	$\mathbf{g}^j = (\eta_j)^{-1}(\mathbf{v}^j - \beta_j \mathbf{g}^{j-1})$
	$\mathbf{x}^j = \mathbf{x}^{j-1} + \xi_j \mathbf{g}^j$, stop if convergence is reached
	$\mathbf{w} = \mathbf{w} - \gamma_j \mathbf{v}^j$
	$\beta_{j+1} = \|\mathbf{w}^j\|$;
	$\mathbf{v}^{j+1} = \mathbf{w}/\beta_{j+1}$
End	

Lanczos algorithm to solve symmetric linear systems

Given:	$\mathbf{x}^{(0)}$
Compute	$\mathbf{r}^{(0)} = \mathbf{b} - A\mathbf{x}^{(0)}$, $\beta = \|\mathbf{r}^{(0)}\|$, $\mathbf{v}^{(1)} = \mathbf{r}^{(0)}/\beta$
Set	$\beta_1 = 0$ and $\mathbf{v}^{(0)} = \mathbf{0}$
For	$k = 1 : m$
	$\mathbf{w}^{(k)} = A\mathbf{v}^{(k)} - \beta_k \mathbf{v}^{(k-1)}$
	$\gamma_k = (\mathbf{w}^{(k)}, \mathbf{v}^{(k)})$
	$\mathbf{w}^{(k)} = \mathbf{w}^{(k)} - \gamma_k \mathbf{v}^{(k)}$
	$\beta_{k+1} = \|\mathbf{w}^{(k)}\|$, if $\beta_{k+1} = 0$, go out of the loop
	$\mathbf{v}^{(k+1)} = \mathbf{w}^{(k)}/\beta_{k+1}$
End	
Set	$T_m = tridiag\{\beta_k, \gamma_k, \beta_{k+1}\}$
Compute	$\mathbf{y}_m = T_m^{-1}(\beta\mathbf{e}^{(1)}) \Longleftarrow \Longleftarrow \Longleftarrow$ Why is that?
	$\mathbf{x}_m = \mathbf{x}_0 + V_m \mathbf{y}_m$

Lanczos algorithm to solve symmetric linear systems

Compute $\mathbf{y}_m = T_m^{-1}(\beta\mathbf{e}^{(1)}) \Longleftarrow \Longleftarrow \Longleftarrow$ Why is that?
$\mathbf{x}_m = \mathbf{x}_0 + V_m \mathbf{y}_m$

Recall: $\mathbf{y} = (W^T A V)^{-1} W^T \mathbf{r}^{(0)}$.
Here $W = V$, $AV^m = V^m H_m$ and $H_m \equiv T_m$.
Thus, $W^T AV = T_m$ and
$W^T\mathbf{r}^{(0)} = V^T\mathbf{r}^{(0)} = \|\mathbf{r}^{(0)}\| V^T \mathbf{v}^{(1)} = \beta V^T \mathbf{v}^{(1)} = \beta \mathbf{e}^{(1)}$,
due to the orthogonality of the columns of V.